Suspect screening and identification of energetic materials using liquid chromatography coupled to high-resolution mass spectrometry

Murphy, Bronagh Ruth

Awarding institution:
King's College London

Suspect screening and identification of energetic materials using liquid chromatography coupled to high-resolution mass spectrometry

Bronagh Ruth Murphy

In partial fulfilment of the requirements for the award of Doctor of Philosophy in

Analytical Science

2019

Supervisors:

Dr. Leon Barron

Dr. Mark Parkin

Department of Analytical, Environmental and Forensic Science,

King's College London

Declaration & Copyright

I confirm that the following thesis does not exceed the word limit prescribed in the College regulations. I further confirm that the work presented in the thesis is my own and all references are cited accordingly.

Bronagh Murphy

April, 2019

Table of Contents

DECLARATION & COPYRIGHT 2

TABLE OF CONTENTS 3

ABSTRACT 6

ACKNOWLEDGEMENTS 8

LIST OF PUBLICATIONS AND CONFERENCE PRESENTATIONS 9

LIST OF FIGURES 11

LIST OF TABLES 18

LIST OF EQUATIONS 20

LIST OF ABBREVIATIONS 21

CHAPTER 1: GENERAL INTRODUCTION 24

1.1 ENERGETIC MATERIALS 25

1.1.1 Explosives legislation in the UK 26

1.1.2 The chemistry of explosives 27

1.1.3 Requirements for forensic analysis of explosives 33

1.2 SCREENING AND IDENTIFICATION OF EXPLOSIVES USING HIGH PERFORMANCE LIQUID CHROMATOGRAPHY COUPLED TO MASS SPECTROMETRY (HPLC-MS) 35

1.2.1 High performance liquid chromatography 35

1.2.2 Mass spectrometry 43

1.2.3 Liquid chromatography - high resolution mass spectrometry (LC-HRMS) 52

1.3 SUSPECT AND NON-TARGET SCREENING 53

1.3.1 Retention time prediction using quantitative structure – retention relationships 54

1.3.2 Evaluating the significance of a match 57

1.4 AIMS AND OBJECTIVES OF THE PHD THESIS 58

1.5 OUTLINE 58

CHAPTER 2: MULTI-RESIDUE LC-HRMS METHOD DEVELOPMENT 60

2.1 INTRODUCTION 61

2.2 EXPERIMENTAL 64
2.2.1 Materials and reagents 64
2.2.2 Optimisation of chromatographic separation conditions 65
2.2.3 Optimisation of high resolution mass spectrometry conditions 65
2.2.4 Optimised LC-HRMS conditions 66
2.2.5 Fingermark deposition and extraction procedures 67
2.2.6 Data analysis 67
2.3 RESULTS AND DISCUSSION 69
2.3.1 Optimisation of chromatography conditions 69
2.3.2 Optimisation of high resolution mass spectrometry 82
2.3.3 Instrumental method performance of optimised multi-residue LC-APCI-HRMS method 91
2.3.4 Application of LC-HRMS method to the detection of explosives in contact traces 92
2.4 CONCLUSION 97

CHAPTER 3: PERFORMANCE OF LC-HRMS METHOD AGAINST AN EXPANDED SET OF COMPOUNDS 99

3.1 INTRODUCTION 100
3.2 EXPERIMENTAL 103
3.2.1 Materials 103
3.2.2 Liquid chromatography-high resolution mass spectrometry conditions 103
3.2.3 Assessing instrumental method performance 103
3.2.4 Passive vapour sampling of an explosives magazine 109
3.2.5 Data analysis 109
3.3 RESULTS AND DISCUSSION 110
3.3.1 HRMS method development 110
3.3.2 Ions detected by LC-(APCI)-HRMS 114
3.3.3 Instrumental method performance 129
3.3.4 Effect of using different identification criteria on selectivity 133
3.3.5 Selectivity in a selection of matrices 144
3.3.6 Detection of fragment ions in fingermark depletion series 147
3.3.7 Method application to passive vapour sampling of explosives 148
3.3.8 Alignment with method validation guidelines 153
3.4 CONCLUSION 156

CHAPTER 4: RETENTION TIME PREDICTION USING MULTIPLE LINEAR REGRESSION AND MACHINE LEARNING 157

4.1 INTRODUCTION 158

4.2 EXPERIMENTAL 159

4.3 RESULTS AND DISCUSSION 163

4.3.1 Building a dataset of measured retention times 163

4.3.2 Developing prediction models with molecular descriptors from previous studies 163

4.3.3 Selecting molecular descriptors for retention time prediction on a C_{18} Ar column 172

4.3.4 Performance of neural networks using selected descriptors 182

4.3.5 Application of retention time prediction to new compounds 190

4.3.6 Prediction of retention times in matrix 207

4.4 CONCLUSION 209

CHAPTER 5: CONCLUSIONS AND FUTURE WORK 211

5.1 SUMMARY 212

5.2 CONCLUSIONS AND RECOMMENDATIONS 217

5.3 RECOMMENDATIONS FOR FUTURE WORK 218

5.3.1 Validation of LC-HRMS method and ANN prediction model 218

5.3.2 Non-target screening 218

5.3.3 Optimisation of sample preparation for suspect screening and non-target analysis 219

5.3.4 LC-MS/HRMS 220

5.3.5 Comprehensive two-dimensional chromatography 221

REFERENCES 223

APPENDIX 231

Abstract

Forensic Scientists need to be able to detect, and confidently identify, any explosive substance encountered during forensic casework. This means flexible analytical methods that provide comprehensive detection of explosives, their precursors, transformation products and related compounds are required. Existing analytical methods typically only target the most commonly encountered explosive substances and often cover relatively small sets of structurally related compounds. Liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) is emerging as a viable technique in many fields for screening, identification and quantification of larger numbers of compounds. The aim of this thesis is to investigate the suitability of LC-HRMS for screening and identification of large numbers of known, unknown and suspect energetic materials.

An LC-HRMS method was developed and the chromatographic separation and ionisation conditions were optimised for a set of eighteen initial target analytes including nitroaromatics, nitramines, nitrate esters and peroxides. The method performance was then assessed for a larger number of target analytes to investigate the potential for suspect screening and non-target analysis of additional energetic materials without further optimisation. The effect of different identification criteria, such as mass accuracy thresholds, number and type of ions and retention time windows, on selectivity and sensitivity was also examined, to support the selection of evidence-based identification requirements for LC-HRMS analysis of energetic materials. Finally, retention time prediction was investigated for the potential to aid preliminary identification of suspect or non-target energetic materials by prioritising acquisition of reference materials or excluding isomers.

Overall, generalisability of a developed LC-HRMS method to a larger set of energetic materials, including MEKP, nitrated sugars and organic gunshot residues, was demonstrated for the first time. This supported the use of full-scan LC-HRMS, for suspect screening and non-target analysis of energetic materials. However, even with high resolution and mass accuracy it was not always possible to unequivocally identify energetic materials using LC-HRMS, due to the presence of hundreds or even thousands of isomers in some cases. In-source fragment ions and the use of ion ratios, increased selectivity but at a cost of sensitivity. The LC-HRMS method was successfully applied to the detection of explosives in contact traces and on passive vapour samplers. The

retention time prediction models and prediction intervals presented in this thesis showed promise for adding value to suspect or non-target screening of energetic materials. Following a critical evaluation of the use of LC-HRMS, alone and in combination with other techniques, good chromatographic separation and tandem mass spectrometry are recommended for confirmatory analysis, along with a consideration of the number of isomers for individual compounds. Due to compromises in selectivity or sensitivity, identification criteria will likely differ for screening versus confirmatory analysis.

Acknowledgements

I would like to acknowledge funding from King's College London, in the form of my GTA PhD Studentship and research funding from Dstl. Without this funding I would not have been able to start, let alone complete, my PhD. I am sincerely grateful to my supervisors for their advice, support and guidance throughout my PhD, from drafting the initial project proposal to providing feedback on a draft of my thesis. I can only apologise for getting somewhat side tracked along the way and the final project being unrecognisable to the original project proposal! Leon, thank you for convincing me the glass is at least half full. Mark, thank you for always being a voice of reason.

I would like to thank the team at the Forensic Explosives Laboratory (FEL) for showing me around your trace lab, donating reference materials and most importantly for providing advice and input, particularly during the first half of my PhD. I would like to thank Gillian McEneff, Dan Wood and Tony Webb for letting me piggy-back on your trials to collect samples for my PhD and providing endless entertainment and good humour. I would also like to everyone at King's Forensics for providing a sounding board, encouragement and light relief throughout. Special thanks to those who got me through the bumps along the way, you know who you are.

Finally, thank you to my parents for your never-ending support, I wouldn't be here without you!

List of Publications and Conference Presentations

Publications

- G L McEneff, **B Murphy,** T Webb, D Wood, R Irlam, J Mills, D Green & L P Barron, Sorbent Film-Coated Passive Samplers for Explosives Vapour Detection Part A: Materials Optimisation and Integration with Analytical Technologies. *Scientific Reports*. 8 (2018)
- G L McEneff, A Richardson, T Webb, D Wood, **B Murphy**, R Irlam, J Mills, D Green & L P Barron, Sorbent Film-Coated Passive Samplers for Explosives Vapour Detection Part B: Deployment in Semi-Operational Environments and Alternative Applications. *Scientific Reports*. 8 (2018)
- M D Gallidabino, L Hamdan, **B Murphy** & L P Barron, Suspect screening of halogenated carboxylic acids in drinking water using ion exchange chromatography – high resolution (Orbitrap) mass spectrometry (IC-HRMS). *TALANTA*. 178 (2018) 57-68.
- H Rapp-Wright, G McEneff, **B Murphy**, S Gamble, R Morgan, M Beardah & L Barron, Suspect screening and quantification of trace organic explosives in wastewater using solid phase extraction and liquid chromatography-high resolution accurate mass spectrometry. *Journal of Hazardous Materials*. 329 (2017) 11-21

Oral Presentations

- **B Murphy**, M Parkin, M Beardah & L Barron, High resolution mass spectrometry: Panacea or Pandora's box for explosives analysis? 12th International Symposium on the Analysis and Detection of Explosives (ISADE). Keble College, Oxford, UK, 17th – 21st September 2017
- **B Murphy**, M Beardah, M Parkin & L Barron, Suspect screening of trace explosives by liquid chromatography high resolution accurate mass spectrometry. ANZFSS 23rd International Symposium on the Forensic Sciences, Auckland, New Zealand, 18th – 23rd September 2016
- **B Murphy**, M Beardah, M Parkin & L Barron, Semi-targeted analysis of organic explosives, precursors and transformation products using liquid chromatography-high

resolution mass spectrometry. Spectrometry for Security Applications Workshop, Birmingham, UK, 17th December 2015

Poster Presentations

- **B Murphy**, M Parkin & L Barron, Optimisation of liquid chromatography-high resolution mass spectrometry (LC-HRMS) for broad screening of organic explosives, precursors and transformation products. 7th European Academy of Forensic Science (EAFS) Conference, Prague, Czech Republic, 6th – 11th September 2015
- **B Murphy**, M Parkin & L Barron, Optimisation of liquid chromatography-high resolution mass spectrometry for semi-targeted screening of explosives, precursors and transformation products. Analytical Research Forum 2015 (ARF2015), London, UK, 3rd July 2015
- **B Murphy**, M Parkin & L Barron, Semi-targeted analytical screening for organic explosives using liquid chromatography-high resolution mass spectrometry and in silico data-mining tools. Forensic International Network for Explosives Investigation (FINEX), Wiesbaden, Germany, 18th – 20th May 2015
- **B Murphy**, M Beardah, M Parkin & L Barron, Development of a multi-residue liquid chromatography-high resolution mass spectrometry screening method for high-order organic explosives detection. Analytical Research Forum 2014 (ARF2014), London, UK, 7th July 2014

List of Figures

Figure 1.1: Classification of energetic materials. Adapted from Singh [2]..... 25

Figure 1.2: van Deemter Plot showing the effect of the A, B and C terms of the van Deemter Equation on plate height with increasing linear velocity..... 36

Figure 1.3: Schematic of negative electrospray ionisation, adapted from [62], [60] and [61]. 47

Figure 1.4: Schematic of positive atmospheric pressure chemical ionisation. Equations outlining the ionisation mechanism are from Awad et al. with S referring to the vaporised solvent and M referring to the analyte [61]. 48

Figure 1.5: Cutaway view of an orbitrap mass analyser. Ions are injected perpendicular to the z-axis, displaced from z=0. This gives the ions potential energy in the z-direction causing them to oscillate along the z-axis. Adapted from Hu et al. 2005 [65]. 50

Figure 1.6: Workflows for three approached to using LC-HRMS: targeted analysis, suspect screening and non-target analysis. Adapted from ([71]). 53

Figure 1.7: Schematic of a feed-forward ANN with an input layer, one hidden layer and an output layer containing four, three and one neurons respectively. Adapted from [73] and [87]. 55

Figure 1.8: Outline of the main advantages, challenges and proposed solutions for the use of LC-HRMS for screening and identification of explosives..... 59

Figure 2.1: Eighteen organic explosives, including sixteen nitro-explosives, grouped based on their chemical structure and identified as initial target compounds for a forensic screen. Note: Tetryl has both nitroaromatic and nitramine functionality . 62

Figure 2.2: Gradient separations of 16 target analytes on a YMC Triart PFP column using: 50-55% B over 0-50 min; 45-55% B over 0-50 min; 40-55% B over 0-50 min and 40-55% B over 0-60 min. All were at 20 °C. 70

Figure 2.3: The effect of temperature on separation of 16 target analytes using the YMC Triart PFP column and a gradient of 42-54 % CH_3OH over 0-50 min. Dashed boxes indicate areas of best separation. 71

Figure 2.4: Optimised separation including two additional target analytes using the YMC Triart PFP column, a gradient of 42-54 % CH_3OH over 0-50 min and a temperature

program of 5-45 °C mid-run. The measured oven temperature throughout the run is indicated by the dashed line. 72

Figure 2.5: Overlaid van Deemter plots, with RDX as a probe, for the comparison of column efficiencies. 74

Figure 2.6: Isocratic separations of 16 target analytes at 210 nm. The position of RDX in each chromatogram is indicated by an asterisk. 76

Figure 2.7: Van't Hoff plot for 16 initial target analytes on a C_{18}Ar column over a temperature range of 5-45 °C. Areas of selectivity change are highlighted in grey, the best separation is indicated by a dashed red rectangle and R^2 values, slope (m) and ΔH values are shown on the right to 3 significant figures. 77

Figure 2.8: Gradient method development on ACE C_{18} Ar column (λ = 210 nm). (i) Shows the effect of different starting mobile phase composition. (ii) Shows multi and singe step gradients starting with 40 % B. (iii) shows the effect of gradient slope with 40-100 % B. 78

Figure 2.9: Optimised separation including some additional target analytes on the ACE C_{18} Ar column using a linear gradient, 40-100 % B over 0-30 min, at 20 °C. 79

Figure 2.10: Scatter plot showing positive correlation between retention factor on the PFP column (K_{PFP}) and retention factor on the C_{18}Ar column (K_{C18AR}) for target analytes, with the orthogonal regression line and equation shown in blue. 81

Figure 2.11: Overlaid LC-(ESI)-HRMS extracted ion chromatograms for target analytes at 0.15, 0.2, 0.25 and 0.3 mL min^{-1}. The total runtime including re-equilibration is also given and indicated by an asterisk. 83

Figure 2.12: Optimisation of ESI-HRMS parameters for the nitro-explosives: a) spray voltage, b) HESI heater temperature and c) capillary temperature. Error bars represent standard deviation of n=3 repeats. Note. 2-NT, 3-NT, 4-NT, 2,6-DNT, NB, EGDN and NG were not detected as deprotonated molecules or adduct ions. 85

Figure 2.13: Average retention times for LC-(ESI)-HRMS with NH_4OAc, NH_4Cl or no additive. Error bars indicate maximum and minimum measured retention times, n=10. . 87

Figure 2.14: Chemical structure and acid dissociation constant (predicted using Percepta PhysChem Profiler) of picric acid, which was predominantly charged at the pH of all mobile phase 88

FIGURE 2.15: OPTIMISATION OF APCI PARAMETERS: A) CORONA DISCHARGE CURRENT, B) VAPORISER TEMPERATURE AND C) CAPILLARY TEMPERATURE. DASHED RED BOXES INDICATE THE SETTING USED FOR OTHER OPTIMISATION EXPERIMENTS. 89

FIGURE 2.16: NUMBER OF INITIAL TARGET ANALYTES DETECTED BY: A) APCI AND B) ESI AT 25, 2.5, 0.5 AND 0.05 NG ON COLUMN. N.B. FOR THE PURPOSE OF THIS FIGURE TETRYL IS CLASSED AS A NITROAROMATIC NOT A NITRAMINE. 91

FIGURE 2.17: EXAMPLE OF SPIKED FINGERMARK WITH AVERAGE RECOVERIES GIVEN IN BRACKETS ± STANDARD DEVIATION (N=6). LC-(APCI)_HRMS EXTRACTED ION CHROMATOGRAMS WITH A 5 PPM MASS ACCURACY FOR A: *M/Z* 60.00910, B: *M/Z* 209.01955, C: *M/Z* 331.01593, D: *M/Z* 61.98837, E: *M/Z* 257.00430, F: *M/Z* 227.98983, G: *M/Z* 123.03258, H: *M/Z* 213.00274, I: *M/Z* 182.03331, J: *M/Z* 136.04040, K: *M/Z* 181.02548, L: *M/Z* 241.02146 AND M: *M/Z* 227.01839. 93

FIGURE 2.18: FINGERMARK DEPLETION SERIES FOLLOWING CONTACT WITH A COMMERCIAL EXPLOSIVE. PART A SHOWS OVERLAID LC-(APCI)-HRMS EXTRACTED ION CHROMATOGRAMS FOR $[RDX+Cl]^-$, *M/Z* 257.00430 ± 5 PPM AND PART B SHOWS OVERLAID EXTRACTED ION CHROMATOGRAM FOR $[HMX+Cl]^-$, *M/Z* 331.01593 ± 5 PPM. DATA POINTS REPRESENT UNSMOOTHED DATA AND LINES ARE PLOTTED WITH 7 POINT GAUSSIAN SMOOTHING. NL = NORMALISED INTENSITY REPRESENTED BY 100%. 95

FIGURE 3.1: LC-HRMS MASS SPECTRA FOR RDX (A), PETN (B) AND TNT (C), WITH FULL-SCAN MS SPECTRA SHOWN IN (I) AND ALL ION FRAGMENTATION WITH HCD COLLISION ENERGY OF 20 SHOWN IN (II). 112

FIGURE 3.2: EFFECT OF ALL ION FRAGMENTATION (AIF) ON EXTRACTED ION CHROMATOGRAMS FOR RDX (A), PETN (B) AND TNT (C) IN A 10 NG μL^{-1} MIXED STANDARD SOLUTION. SOLID LINES REPRESENT FULL-SCAN MS AND DASHED LINES INDICATE AIF WITH A HCD CELL COLLISION ENERGY OF 20. 113

FIGURE 3.3: SIGNAL-TO-NOISE RATIO OF THE IONS IDENTIFIED AS ION 1, 2 AND 3 FOR EACH COMPOUND IN A 10 NG μL^{-1} STANDARD (N=60, HMDD AND MEKP OLIGOMERS NOT INCLUDED). 118

FIGURE 3.4: LOSS OF HYDROXYL RADICAL FROM ORTHO-ACI FORM OF TNT RADICAL ANION. ADAPTED FROM [113]. 120

FIGURE 3.5: CHROMATOGRAPHIC SEPARATION OF: A) MANNITOL HEXANITRATE AND B) SORBITOL HEXANITRATE FROM THE LESS NITRATED SUGARS, PENTANITRATE (MPH AND SPN),

tetranitrate (MTN and STN) trinitrate (MTriN ansSTriN) and dinitrate (MDN and SDN). 122
Figure 3.6: Chemical structures of A) mannitol hexanitrate, B) sorbitol hexanitrate and C) the three isomers of mannitol pentanitrate. 123
Figure 3.7: Mechanisms proposed by Colizza et al. for A) the addition of one molecule of methanol to TATP and B) the addition of a second methanol molecule [116]. 125
Figure 3.8: Proposed mechanism for the addition of one methanol molecule to DADP. . 126
Figure 3.9: Chemical structures and monoisotopic masses for cyclic peroxide (CP) and dihydroperoxy peroxide (DHP) oligomers of MEKP, which were detected as ammonium adducts. 126
Figure 3.10: Extracted ion chromatograms showing the presence of multiple oligomers in a solution of MEKP. 127
Figure 3.11: Pie charts summarising the number of elemental compositions (EC) within 5, 2 or 1 ppm of the exact mass of each compound (n=60, MEKP oligomers and ammonium picrate not included). 136
Figure 3.12: Boxplot showing the effect of mass accuracy thresholds and including/excluding chlorine on the number of elemental compositions (n=60, MEKP oligomers and ammonium picrate not included). Boxes indicate interquartile range, whiskers are 1.5 times the height of the box, circles show outliers with values 1.5-3 times height of box and stars show extreme outliers with values >3 times the height of the box. 136
Figure 3.13: Correlation between monoisotopic mass and number of molecular formulae within 1, 2 and 5 ppm (n=60, MEKP oligomers and ammonium picrate not included) ... 137
Figure 3.14: Mass accuracy of the three ions detected for each compound in a 10 ng μL^{-1} standard (n=61, MEKP oligomers not included). 138
Figure 3.16: Pie chart summarising the number of matches from a molecular formula search of the ChemSpider database for each compound (n=60, MEKP oligomers and ammonium picrate not included). 140
Figure 3.16: Selectivity achieved between explosive compounds when HRMS ions were used for identification without retention time windows (n=60, MEKP oligomers and HMDD not included). “M type” = molecular ion, (de)protonated molecule or adduct ion. 141

FIGURE 3.17: A COMPARISON OF THE NUMBER OF EXPLOSIVE RELATED COMPOUNDS RESULTING IN FALSE POSITIVES (WITHIN 5 PPM) WITH AND WITHOUT THE USE OF RETENTION TIME WINDOWS FOR IDENTIFICATION OF COMPOUNDS WITH A RETENTION FACTOR (K) GREATER THAN 1 (N=47). 142

FIGURE 3.18: ION RATIOS FOR 2,4-, 2,6- AND 3,4-DINITROTOLUENE ISOMERS. ERROR BARS REPRESENT 95 % CONFIDENCE INTERVALS, N=6. 144

FIGURE 3.19: SELECTIVITY ISSUES IN 18 MATRICES WHEN RELYING ON HRMS IONS (*M/Z* ± 5 PPM) ONLY FOR IDENTIFICATION OF TARGET COMPOUNDS (N=61, MEKP OLIGOMERS NOT INCLUDED). 145

FIGURE 3.20: SELECTIVITY ISSUES IN 18 MATRICES WHEN BOTH HRMS IONS AND RETENTION WINDOWS (*M/Z* ± 5 PPM, T_R ± 2.5 %, K > 1) WERE USED FOR IDENTIFICATION OF TARGET COMPOUNDS (N=47). 145

FIGURE 3.21: FINGERMARK DEPLETION SERIES EICS FOR RDX ION 1 (257.0043 ± 5 PPM, DARK BLUE), RDX ION 2 (102.0309 ± 5 PPM, MID BLUE) AND RDX ION 3 (129.0418 ± 5 PPM, LIGHT BLUE). A 0.01 NG μL^{-1} RDX STANDARD IS SHOWN IN A; THE 1ST, 11TH, 21ST, 31ST, 41ST, 51ST, AND 61ST FINGERMARK OF THE DEPLETION SERIES ARE SHOWN IN B-H AND I SHOWS A BLANK FINGERMARK. LINES SHOW 7-POINT GAUSSIAN SMOOTHING. 146

FIGURE 3.22: ION RATIOS FOR RDX FRAGMENTS VERSUS CHLORIDE ADDUCT IONS IN FINGERMARK DEPLETION SERIES. LINES SHOW AVERAGE ION RATIO IN A 1 NG μL^{-1} RDX STANDARD (N=6) AND ± 30 %, THE PERMITTED TOLERANCE FOR IONS WITH RELATIVE INTENSITY OF 10-20 % ACCORDING TO 2002/657/EC [93]. 147

FIGURE 3.23: PASSIVE VAPOUR SAMPLING OF AN EXPLOSIVES MAGAZINE WITH OVERLAID EXTRACTED ION CHROMATOGRAMS SHOWING IONS 1-3 FOR: A) 1,2- AND 1,3-DNB; B) 2,4-, 2,6- AND 3,4-DNT; C) TNT; D) 2- AND 4-NT; E) RDX; F) DMNB; G) DPA AND H) EGDN..... 149

FIGURE 3.24: ION RATIOS FOR IONS DETECTED ON PASSIVE SAMPLERS. BLUE LINES SHOW AVERAGE ION RATIO IN STANDARD (N=6) AND GREEN LINES SHOW UPPER AND LOWER THRESHOLDS..... 152

FIGURE 4.1: CORRELATION BETWEEN MEASURED AND PREDICTED RETENTION TIMES (A) AND PREDICTION ERROR (B) FOR THE BEST MODELS USING THREE DIFFERENT SETS OF MOLECULAR DESCRIPTORS. GREY LINES INDICATE PERFECT PREDICTION AND DASHED GREEN LINES INDICATE PERCENTILES FOR THE COMPLETE DATASET..... 165

FIGURE 4.2: PERFORMANCE OF PREVIOUS MLR MODELS ON THIS DATASET WITH A) SHOWING THE CORRELATION BETWEEN PREDICTED AND MEASURED RETENTION TIMES AND B) SHOWING THE PREDICTION ERROR ACROSS THE RUN. GREY LINES INDICATE PERFECT PREDICTION AND DASHED GREEN LINES SHOW PERCENTILES. 169

FIGURE 4.3: PERFORMANCE OF MLR MODELS DEVELOPED FOR THIS DATASET WITH A) SHOWING CORRELATION BETWEEN PREDICTED AND MEASURED RETENTION TIMES AND B) SHOWING THE PREDICTION ERROR OVER TIME. GREY LINES INDICATE PERFECT PREDICTION. 171

FIGURE 4.4: COLLINEARITY BETWEEN A) BLTD48 AND BLTA96 AND B) BLTD48 AND MLOGP. LINES OF BEST FIT ARE SHOWN IN DARK BLUE FOR EXPLOSIVES, LIGHT BLUE FOR DRUGS AND BLACK FOR THE COMBINED DATASET. 174

FIGURE 4.5: LINEAR CORRELATION BETWEEN MEASURED RETENTION TIME AND A) ACD/LOGD AT PH 7, B) ACD/LOGP C) BLTD48 AND D) NUMBER OF CARBON ATOMS (NC). LINES OF BEST FIT ARE SHOWN IN DARK BLUE FOR EXPLOSIVES, LIGHT BLUE FOR DRUGS AND BLACK FOR THE COMBINED DATASET. 175

FIGURE 4.6: CORRELATION BETWEEN LOGP AND LOGD AT PH 7 FOR A) EXPLOSIVES AND B) DRUGS WITH THE DASHED LINE SHOWING X=Y. 176

FIGURE 4.7: CORRELATION BETWEEN MEASURED AND PREDICTED RETENTION TIMES (A) AND PREDICTION ERROR (B) FOR THE BEST MODELS USING A SUBSET OF DESCRIPTORS SELECTED BY (I) FORWARD SELECTION AND (II) GENETIC ALGORITHM. GREY LINES INDICATE PERFECT PREDICTION AND DASHED GREEN LINES SHOW PERCENTILES. 183

FIGURE 4.8: CORRELATION BETWEEN MEASURED AND PREDICTED RETENTION TIMES (A) AND PREDICTION ERROR (B) FOR THE ENSEMBLES (N=6) USING A SUBSET OF DESCRIPTORS SELECTED BY (I) FORWARD SELECTION AND (II) GENETIC ALGORITHM. GREY LINES INDICATE PERFECT PREDICTION AND DASHED GREEN LINES SHOW PERCENTILES. 185

FIGURE 4.9: RANK ORDER CORRELATION FOR A) ALL 149 COMPOUNDS AND B) NITRATE ESTERS USING I) THE ENSEMBLE OF 6 MLP:32-2-1 AND II) THE BEST MLP:11-3-1. 187

FIGURE 4.10: ASSESSMENT OF NORMAL DISTRIBUTION OF PREDICTION ERRORS USING A) HISTOGRAMS AND B) NORMAL Q-Q PLOTS FOR I) 6XMLP:33-2-1 AND II) MLP:11-3-1. 189

FIGURE 4.11: EICS FROM GUN SURVEILLANCE STANDARD FOR A) NDPAS, B) DNDPAS, C) DMP AND D) N-NDPA. 190

FIGURE 4.12: CHEMICAL STRUCTURES OF NEW OGSR COMPOUNDS USED AS AN EXTERNAL TEST SET FOR VALIDATION OF PREDICTION MODELS. 191

FIGURE 4.13: FORMATION OF A $C_6H_9O_{14}N_4^+$ ION FROM MANNITOL HEXANITRATE (MHN), PENTANITRATE (MPN) AND TETRANITRATE (MTN). 195

FIGURE 4.14: EXPERIMENTAL AND PREDICTED RETENTION TIMES FOR RDX, TATB AND PYX COMPARED TO PREDICTED RETENTION TIMES OF STRUCTURAL ISOMERS USING A) ENSEMBLE OF

6xMLP:32-2-1 and B) MLP:11-3-1. Grey error bars show 99 % PI and black error bars show 75 % PI. 206
Figure 4.15: Effect of pond water and fingermark matrices on retention times for drugs and explosives. Vertical error bars indicate standard deviation in matrix (n=3 for pond water, n=6 for fingermarks) and x-error bars indicate standard deviation in solvent (n=6). Dashed grey lines show ± 2.5 % retention time in solvent. 208

List of Tables

TABLE 1.1: SUMMARY OF MAIN OFFENCES RELATING TO EXPLOSIVES. ADAPTED FROM [3]. 26

TABLE 1.2: EXAMPLES OF NITRAMINE, NITRATE ESTER, NITROAROMATIC AND PEROXIDE EXPLOSIVES. 28

TABLE 1.3: ORGANIC EXPLOSIVE PRECURSORS AND INTERMEDIATES 30

TABLE 1.4: EXAMPLES OF EXPLOSIVE STABILISERS AND PLASTICISERS 32

TABLE 1.5: LIQUID CHROMATOGRAPHY METHODS FOR THE SEPARATION OF ORGANIC EXPLOSIVES..... 38

TABLE 1.6: DETECTION DETAILS FOR LC-MS ANALYSIS OF MULTIPLE EXPLOSIVES 44

TABLE 1.7: PREDICTION OF LC RETENTION TIME FOR NEW COMPOUNDS USING ANNS. 57

TABLE 2.1: DETAILS OF COLUMNS TESTED FOR EFFICIENCY AND SELECTIVITY. 65

TABLE 2.2: RETENTION, SYMMETRY, SELECTIVITY AND RESOLUTION OF OPTIMISED SEPARATION ON THE YMC TRIART PFP COLUMN. 73

TABLE 2.3: RETENTION, SYMMETRY, SELECTIVITY AND RESOLUTION OF OPTIMISED SEPARATION ON THE ACE C_{18} AR COLUMN. 79

TABLE 2.4: LC-(ESI)-HRMS SENSITIVITY AND LINEAR RANGE FOR TARGET NITRO-EXPLOSIVES DETECTED IN NEGATIVE MODE USING AMMONIUM CHLORIDE, AMMONIUM ACETATE OR NO MOBILE PHASE ADDITIVE. 86

TABLE 2.5: APCI METHOD PERFORMANCE FOR EIGHTEEN INITIAL TARGET ANALYTES. 92

TABLE 2.6: COMPOUNDS DETECTED BY LC-(APCI)-HRMS IN FINGERMARK DEPLETION SERIES USING DEFAULT PEAK DETECTION SETTINGS WITH AND WITHOUT SMOOTHING. 96

TABLE 3.1: PURITY, CONCENTRATION AND SOURCE OF REFERENCE MATERIALS. 104

TABLE 3.2: IONS DETECTED BY LC-(APCI)-HRMS METHOD INCLUDING PROPOSED IDENTITY AND MASS ACCURACY (N=61 COMPOUNDS, MEKP OLIGOMERS NOT INCLUDED). 115

TABLE 3.3: LC-APCI-HRMS PERFORMANCE: ESTIMATED LIMITS OF DETECTION (LOD), LINEARITY, REPEATABILITY, REPRODUCIBILITY AND DETECTION PROBABILITY (N= 61, MEKP OLIGOMERS NOT INCLUDED) 130

TABLE 3.4: ISOMERS AND ELEMENTAL COMPOSITIONS WITHIN 5, 2 AND 1 PPM OF MONOISOTOPIC MASS (N=60 COMPOUNDS, MEKP OLIGOMERS AND AMMONIUM PICRATE NOT INCLUDED) 134

TABLE 4.1: GROUPS OF MOLECULAR DESCRIPTORS GENERATED USING PARAMETER CLIENT [133].. 161

TABLE 4.2: MOLECULAR DESCRIPTORS USED IN PREVIOUS RETENTION TIME PREDICTION MODELS ... 164

TABLE 4.3: PERFORMANCE OF THE BEST ANN FOR EACH SET OF DESCRIPTORS 167

TABLE 4.4: PERFORMANCE OF PUBLISHED MLR MODELS ON THIS DATASET. 170

TABLE 4.5: PERFORMANCE OF MLR MODELS OPTIMISED FOR THIS DATASET. 171

TABLE 4.6: SUBSET OF 32 FORWARD SELECTED DESCRIPTORS AND SENSITIVITY OF MLP:32-2-1 TO THESE DESCRIPTORS 179

TABLE 4.7: SUBSET OF 11 GENETIC ALGORITHM DESCRIPTORS AND SENSITIVITY OF MLP:11-3-1 TO THESE DESCRIPTORS 182

TABLE 4.8: PERFORMANCE OF THE BEST MODEL AND AN ENSEMBLE OF 6 MODELS USING A SUBSET OF 32 OF THE MOLECULAR DESCRIPTORS SELECTED BY FORWARD SELECTION 183

TABLE 4.9: PERFORMANCE OF BEST MODEL AND AN ENSEMBLE OF SIX MODELS USING A SUBSET OF THE DESCRIPTORS SELECTED BY GENETIC SELECTION 184

TABLE 4.10: SPEARMAN'S RANK ORDER CORRELATION BETWEEN PREDICTED AND MEASURED RETENTION TIMES 186

TABLE 4.11: PREDICTION INTERVALS 188

TABLE 4.12: MEASURED AND PREDICTED RETENTION TIMES FOR OGSR COMPONENTS. 192

TABLE 4.13: PREDICTED RETENTION TIME AND ORDER FOR MEKP OLIGOMERS USING 6XMLP:32-2-1. 193

TABLE 4.14: LOGP VALUES AND ORDER FOR MEKP OLIGOMERS 194

TABLE 4.15: PREDICTED RETENTION TIMES FOR MANNITOL NITRATES USING ENSEMBLE OF 6XMLP:32-2-1 196

TABLE 4.16: PREDICTED RETENTION TIMES FOR SORBITOL NITRATES USING THE ENSEMBLE OF 6XMLP:32-2-1 198

TABLE 4.17: IONS DETECTED FOR EACH CHROMATOGRAPHIC PEAK IN SHN SAMPLE 201

TABLE 4.18: PREDICTED RETENTION TIMES FOR COMPOUNDS WITH THE SAME ELEMENTAL COMPOSITION AS RDX ($C_3H_6O_6N_6$), PYX ($C_3H_6O_6N_6$) AND TATB ($C_6H_6O_6N_6$). 202

List of Equations

EQUATION 1-1: RESOLUTION 36

EQUATION 1-2: EFFICIENCY 36

EQUATION 1-3: VAN DEEMTER EQUATION 36

EQUATION 2-1: RETENTION FACTOR 68

EQUATION 2-2: SELECTIVITY FACTOR 68

EQUATION 2-3: RESOLUTION USING PEAK WIDTH AT HALF HEIGHT 68

EQUATION 2-4: VAN'T HOFF EQUATION 70

EQUATION 3-1: LIMIT OF DETECTION 107

EQUATION 4-1: PREDICTION INTERVAL 188

EQUATION 4-2: STANDARD ERROR OF PREDICTION 188

List of Abbreviations

Abbreviation	Meaning
A	Selectivity factor
A-DNT	Amino-dinitrotoluene
AIF	All ion fragmentation
ANN	Artificial neural network
APCI	Atmospheric pressure chemical ionisation
API	Atmospheric pressure ionisation
Ar	Aromatic
A_s	Peak symmetry
Aspirin	Acetyl salicylic acid
C4	Composition 4
Cal 1	Lowest concentration standard
CI	Confidence interval
CID	Collision induced dissociation
CP	Cyclic peroxide
DADP	Diacetone diperoxide
DA-NT	Diamino-nitrotoluene
DAP	Diaminopropane
DDNP	Diazodinitrophenol
DEDPU	Diethyl diphenylurea
DEGDN	Diethylene glycol dinitrate
DHP	Dihydroperoxy peroxide
DMDPU	Dimethyl diphenylurea
DMNB	Dimethyl dinitrobutane
DMP	Dimethyl phthalate
DNA	Dinitroaniline
DNB	Dinitrobenzene
DNDPA	Dinitrodiphenylamine
DNG	Dinitroglycerin
DNSA	Dinitrosalicylic acid
DNT	Dinitrotoluene
DPA	Diphenylamine
EC	Elemental composition
EGDN	Ethylene glycol dinitrate
EIC	Extracted ion chromatogram
ENFSI	European Network of Forensic Science Institutes
EPA	Environmental Protection Agency
ESI	Electrospray ionisation
ETN	Erythritol tetranitrate
FWHM	Full width at half maximum peak height
GC	Gas chromatography
GC/MS	Gas chromatography/mass spectrometry
GRNN	Generalised regression neural network
H	Plate height
HCD	Higher-energy C-trap dissociation
HESI	Heated electrospray ionisation
HMDD	Hexamethylene diperoxide diamine
HMTA	Hexamethylenetetramine
HMTD	Hexamethylene triperoxide diamine
HMX	High melting point explosive
HNDPA	Hexanitrodiphenylamine
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IP	Identification point
IUPAC	International Union of Pure and Applied Chemistry

Abbreviation	Meaning
K	Retention factor
LC	Liquid chromatography
LC-HRMS	Liquid chromatography – high resolution mass spectrometry
LC-MS	Liquid chromatography – mass spectrometry
LOD	Limit of detection
MAE	Mean absolute error
MANOVA	Multivariate analysis of variance
MBE	Mean bias error
MDN	Mannitol dinitrate
MEKP	Methyl ethyl ketone peroxide
MHN	Mannitol hexanitrate
MLP	Multilayer perceptron
MLR	Multiple linear regression
MNG	Mononitroglycerin
MPN	Mannitol pentanitrate
MRFA	Met-Arg-Phe-Ala
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MTN	Mannitol tetranitrate
MTriN	Mannitol trinitrate
N	Efficiency
NB	Nitrobenzene
NDPA	Nitrodiphenylamine
NG	Nitroglycerin
NICI	Negative ion chemical ionisation
NL	Maximum normalised intensity
NM	Nitromethane
N-NDPA	N-nitrosodiphenylamine
NQ	Nitroguanidine
NT	Nitrotoluene
PA	Picric acid
PETN	Pentaerythritol tetranitrate
PFP	Pentafluorophenyl
PGDN	Propylene glycol dinitrate
PI	Prediction interval
PNN	Probabilistic neural network
PYX	2,6-Bis(Picrylamino)-3,5-dinitropyridine
QSRR	Quantitative structure-retention relationships
QTOF	Quadrupole time-of-flight
R	Pearson correlation coefficient
RBF	Radial basis function
RDX	Research department explosive
RMSE	Root mean squared error
R_s	Resolution
R-Salt	1,3,5-trinitroso-1,3,5-triazinane
RSD	Relative standard deviation
SD	Standard deviation
SDN	Sorbitol dinitrate
SHN	Sorbitol hexanitrate
SMILES	Simplified Molecular Input Line Entry Specification
SMN	Sorbitol mononitrate
SPE	Solid phase extraction
SPN	Sorbitol pentanitrate
STN	Sorbitol tetranitrate
STriN	Sorbitol trinitrate
t_0	Void time
TATB	Triaminotrinitrobenzene
TATP	Triacetone triperoxide
TEGDN	Triethylene glycol dinitrate
Tetryl	N-methyl-N-(2,4,6-trinitrophenyl)nitramide

Abbreviation	Meaning
TMETN	Trimethylolethane trinitrate
TNB	Trinitrobenzene
TNT	Trinitrotoluene
t_R	Retention time
UHPLC	Ultra-high performance liquid chromatography
UV	Ultra-violet
W	Peak width

Chapter 1: General introduction

1.1 Energetic Materials

Energetic materials contain a large amount of energy that can be released in powerful exothermic combustion reactions, where heated gaseous reaction products rapidly expand. This expansion of hot gases can be used to propel a projectile (e.g. ammunition), or to produce a destructive pressure wave (i.e. an explosion). Energetic materials can be classified as either high explosives which detonate, producing a pressure wave that travels faster than the speed of sound, or low explosives which deflagrate, producing a pressure wave that travels slower than the speed of sound (Figure 1.1). High explosives can be further classified as either primary or secondary explosives based on their sensitivity to detonation. As primary explosives are more sensitive to detonation, small amounts of primary explosive are used to detonate less sensitive secondary explosives which are used as the main charge for military and commercial purposes, due to the comparative ease of handling, storage and transportation. The very high sensitivity of triacetone triperoxide (TATP) and hexamethylene triperoxide diamine (HMTD) to impact, friction and heat means they are considered too high risk for military or commercial purposes, but this has not prevented their use as homemade explosives in improvised devices. Some authors further split high explosives into primary, secondary and tertiary explosives [1]. This reflects the use of three stages in an explosive train, with a detonator, followed by a booster and then the main charge, in some instances.

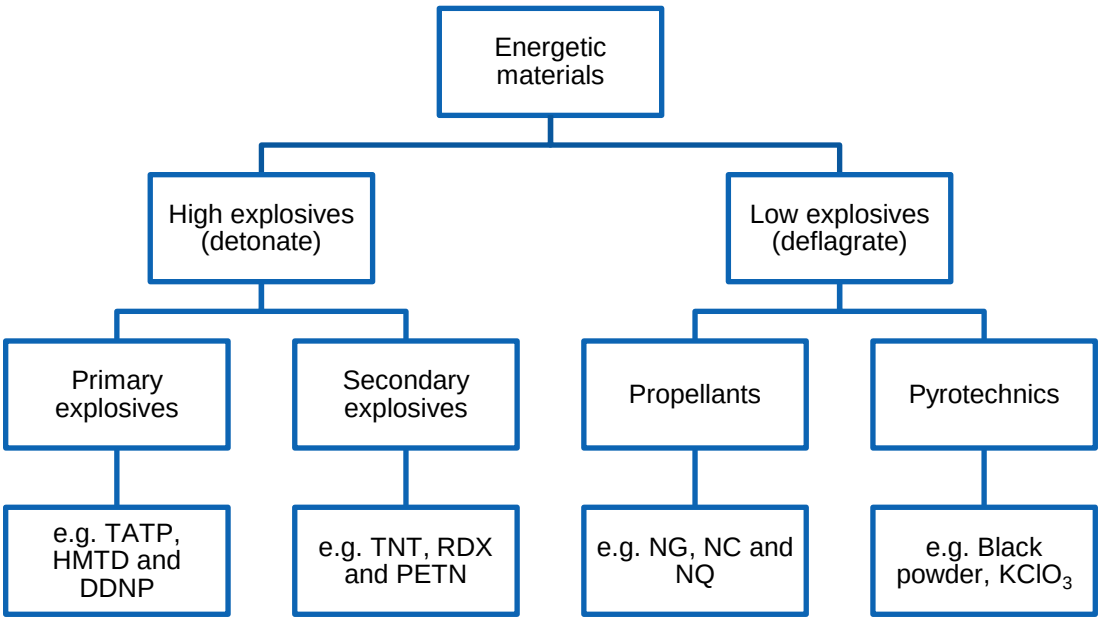


Figure 1.1: Classification of energetic materials. Adapted from Singh [2]

1.1.1 Explosives legislation in the UK

When the manufacture, possession or use of an explosive substance is alleged, forensic scientists are asked to provide evidence to identify the substances in question and, for offences relating to an explosion, provide evidence that an explosion has occurred [3]. Additionally, for offences regarding explosive precursors and/or the manufacture of explosives, forensic scientists may also be required to identify explosive precursors and intermediates. A summary of the main offences, and the relevant legislation, relating to explosives is provided in Table 1.1. There are several pieces of legislation regarding explosives including the: Offences Against the Person Act (OAPA) 1861 [4], Criminal Damage Act 1971 [5], Poisons Act 1972 [6,7], Explosives Act 1875 [8], Criminal Law Act 1977 [9], Explosives Substances Act (ESA) 1883 [10], Terrorism Act 2000 [11], Fireworks Regulations 2004 [12] and Explosives Regulations 2014 [13,14]. The definition of an explosive substance varies between the different pieces of legislation. According to the Explosives Regulations 2014 an 'explosive substance' is a non-gaseous substance or preparation which is:

> "(a) capable by chemical reaction in itself of producing gas at such a temperature and pressure and at such a speed as could cause damage to surroundings; or
>
> (b) designed to produce an effect by heat, light, sound, gas or smoke, or a combination of these as a result of a non-detonative, self-sustaining, exothermic chemical reaction;" [13].

The definition of an 'explosive substance' included in the Explosive Substances Act 1883 also includes "any materials for making any explosive substance" [10]. Therefore, forensic scientists need to be able to identify any substance that can either produce an explosion or be used to make another substance that can produce an explosion. Given that there are thousands of known explosive substances [15], this is a challenging task.

Table 1.1: Summary of main offences relating to explosives. Adapted from [3].

Offence	Relevant legislation	Maximum sentence
Cause an explosion likely to endanger life or cause serious injury to property	ESA 1883, Section 2	Life
Do any act with intent to cause, or conspiring to cause, an explosion likely to endanger life or cause serious injury to property	ESA 1883, Section 3(1) a	Life
Makes, possesses or controls an explosive substance with intent to endanger life or cause serious injury to property	ESA 1883, Section 3(1) b	Life
Makes, knowingly has in control or possession an explosive in suspicious circumstances	ESA 1883, Section 4(1)	Life
Bomb making documents and/or recipes for the production of explosives (including pyrotechnics- low explosives)	Terrorism Act 2000, Section 57 and 58	15 years and/or a fine
Cause grievous bodily harm by the unlawful and malicious explosion of gunpowder or other explosive substance	OAPA 1861, Section 28	Life

Table 1.1 (Continued): Summary of main offences relating to explosives. Adapted from [3].

Offence	Relevant legislation	Maximum sentence
Cause gunpowder or some other explosive substance to explode with the intent to cause grievous bodily harm	OAPA 1861, Section 29	Life
Placing explosives near buildings or ships with intent to do bodily injury	OAPA 1861, Section 30	14 years
Make or have an explosive substance with intent to commit any felony against the Act	OAPA 1861, Section 64	2 years
Causing or intending to cause damage or destroying property	Criminal Damage Act 1971, Section 1(1) and (2)	Life
Place or post any article with the intention of inducing someone to believe that it is likely to ignite or explode and cause injury or damage property	Criminal Law Act 1977, Section 51(1)	7 years
Possession, of listed explosives precursors without a licence - after 3 March 2016	Poisons Act 1972, Section 3(1). Amended by Deregulation Act 2015, Schedule 21	2 years.
Selling of listed explosives precursors to a person who does not hold a licence	Poisons Act 1972, Section 3A. Amended by Deregulation Act 2015, Schedule 22	2 years.
Failing to report a suspicious transaction, loss or theft of a regulated or reportable precursor	Poisons Act 1972, Section 3C. Amended by Deregulation Act 2015, Schedule 23	3 months and/or a fine
Acquiring /keeping explosives without a valid certificate*	Explosives Regulations 2014, Regulation 5 (1) and (2).	2 years and/or fine**
Manufacturing more than 100g of explosives without a licence	Explosives Regulations 2014, Regulation 6 (1) and (2).	2 years and/or fine**
Possession of an F2, F3 or F4 firework by a person under 18 in a public place	Fireworks Regulations, 2004 (SI 1836: 2004), Regulation 4	6 months and/or fine
Any person to be in possession of category 4 fireworks	Fireworks Regulations, 2004 (SI 1836: 2004), Regulation 5	6 months and/or fine
Using a firework at night other than during a permitted fireworks night	Fireworks Regulations, 2004 (SI 1836: 2004), Regulation 7	6 months and/or fine
Throwing or discharging a firework in a public place	The Explosives Act 1875, Section 80	Fine

* The explosive solutions analysed as part of this PhD were exempt from the Explosives Regulations 2014 according to Schedule 2, Part 2, 8(a) on the grounds that they contain less than 5 g of desensitised explosives and were acquired for research and analysis at a University.
** Information obtained from [16].

1.1.2 The chemistry of explosives

In addition to the classification scheme shown in Figure 1.1, energetic materials can also be divided into organic and inorganic compounds and classified based on the functional groups they contain. For combustion, a fuel and an oxidant are required in appropriate quantities and concentrations. In the case of organic explosives, the focus of this thesis, the fuel and oxidant are contained within the same molecule; although depending on the oxygen balance of the molecule, additional oxygen may also be required [17]. The hydrocarbon backbone of organic explosives provides the fuel and functional groups containing oxygen and nitrogen act as the oxidants. Examples of some of the main chemical classes of organic explosives, encountered in forensic investigations: nitramines (R-N-NO_2), nitrate esters (R-O-NO_2), nitroaromatics (Ar-NO_2) and peroxides (R-O-O-R), are shown in Table 1.2.

Table 1.2: Examples of nitramine, nitrate ester, nitroaromatic and peroxide explosives.

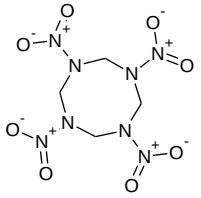
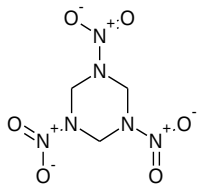
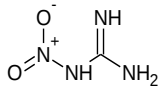
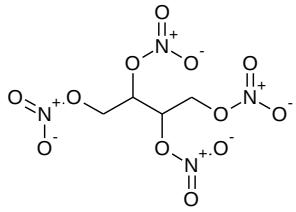
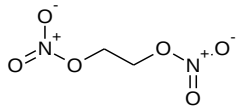
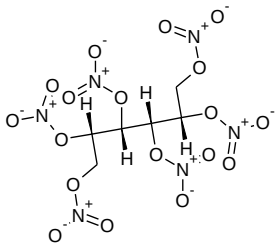
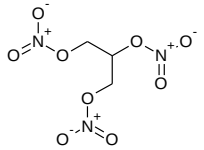
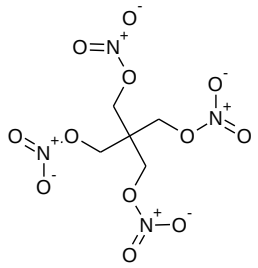
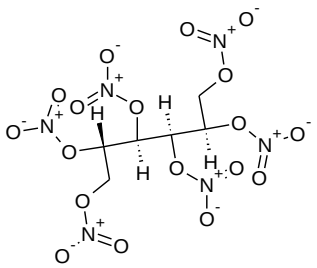
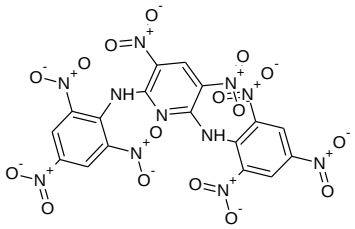
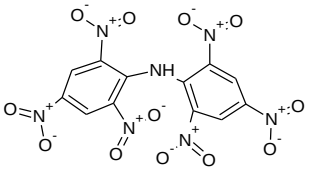
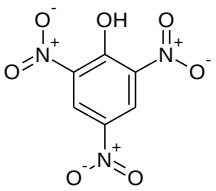
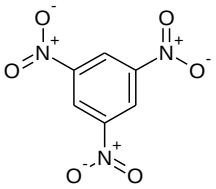
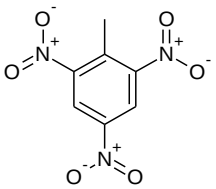
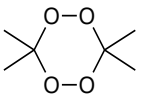
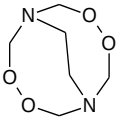
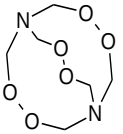
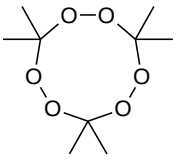
Name	Structure	Class
1,3,5,7-Tetranitro-1,3,5,7-tetrazocane (HMX)		Nitramine
1,3,5-Trinitro-1,3,5-triazine (RDX)		Nitramine
Nitroguanidine (NQ)		Nitramine
Erythritol tetranitrate (ETN)		Nitrate ester
Ethylene glycol dinitrate (EGDN)		Nitrate ester
Mannitol hexanitrate (MHN)		Nitrate ester
Nitroglycerin (NG)		Nitrate ester
Pentaerythritol tetranitrate (PETN)		Nitrate ester
Sorbitol hexanitrate (SHN)		Nitrate ester

Table 1.2 (Continued): Examples of nitramine, nitrate ester, nitroaromatic and peroxide explosives.

Name	Structure	Class
1,3,5-Triamino-2,4,6-trinitrobenzene (TATB)		Nitroaromatic
3,5-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-2,6-pyridinediamine (PYX)		Nitroaromatic
Hexanitrodiphenylamine (HNDPA)		Nitroaromatic
Picric acid (PA)		Nitroaromatic
Trinitrobenzene (TNB)		Nitroaromatic
Trinitrotoluene (TNT)		Nitroaromatic
Diacetone diperoxide (DADP)		Peroxide
Hexamethylene diperoxide diamine (HMDD)		Peroxide
Hexamethylene triperoxide diamine (HMTD)		Peroxide
Triacetone triperoxide (TATP)		Peroxide

1.1.2.1 Explosive precursors and intermediates

In the case of homemade explosives (HME), the availability of precursors can influence the type of explosives synthesized. For example, erythritol tetranitrate (ETN) has seen an increased use following the wider availability of precursor erythritol, which is now commonly used as an artificial sweetener [18,19]. It is for this reason that eight explosive precursors (including nitric and sulfuric acid which are used for nitration reactions) are regulated in the UK and require an Explosives Precursors and Poisons (EPP) license to purchase or supply them above a threshold concentration (e.g. 3 % w/w for nitric acid and 15 % w/w for sulphuric acid) [6,7]. A further 10 explosive precursors are listed as reportable substances. Failure to report a suspicious transaction, loss or theft of a reportable substance is also an offence [6,7]. The lists of explosive precursors included in the Poisons Act 1972 and the Poisons Act 1972 (Explosives Precursors) (Amendment) Regulations 2018 are by no means comprehensive lists of explosive precursors. Other precursors and intermediates are also of interest in forensic science, especially for providing evidence that explosives have been/could be synthesized at a scene. Some examples of organic precursors and intermediates of interest to forensic science are shown in Table 1.3. In some cases, such as with the DNT isomers, precursors and intermediates may also be energetic themselves. As well as a precursor to TNT, 4-NT is one of three compounds (along with ethylene glycol dinitrate (EGDN) and 2,3-dimethyl-2,3-dinitrobutane (DMNB)) that can be used as a detection agent. A detection agent must be added to all plastic explosives, according to The Explosives Regulations 2014 [13].

Table 1.3: Organic explosive precursors and intermediates

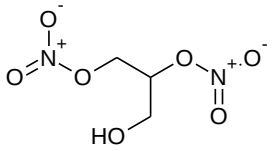
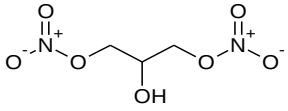
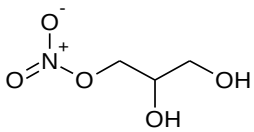
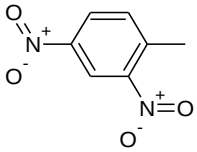
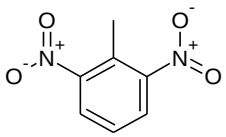
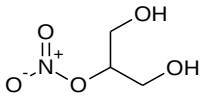
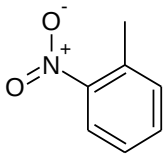
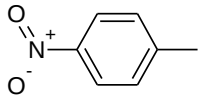
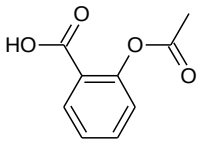
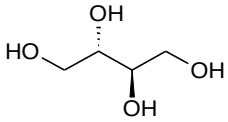
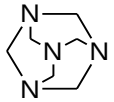
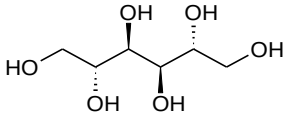
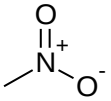
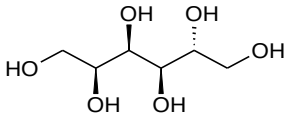
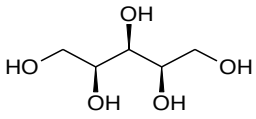
Precursor/Intermediate	Structure	Explosive product
1,2-Dinitroglycerin (1,2-DNG)		Nitroglycerin (NG)
1,3-Dinitroglycerin (1,3-DNG)		Nitroglycerin (NG)
1-Nitroglycerin (1-MNG)		Nitroglycerin (NG)
2,4-Dinitrotoluene (2,4-DNT)		Trinitrotoluene (TNT)

Table 1.3 (Continued): Organic explosive precursors and intermediates

Precursor/Intermediate	Structure	Explosive product
2,6-Dinitrotoluene (2,6-DNT)		Trinitrotoluene (TNT)
2-Nitroglycerin (2-MNG)		Nitroglycerin (NG)
2-Nitrotoluene (2-NT)		Trinitrotoluene (TNT)
4-Nitrotoluene (4-NT)		Trinitrotoluene (TNT)
Acetone[a]		Triacetone triperoxide (TATP)
Acetylsalicylic acid (Aspirin)		Picric acid (PA)
Erythritol		Erythritol tetranitrate (ETN)
Hexamethylenetetramine (HMTA)[a,b]		HMTD, HMX or RDX
Mannitol		Mannitol hexanitrate (MHN)
Nitromethane (NM)[c]		Ammonium nitrate-nitromethane mix (AN/NM)
Sorbitol		Sorbitol hexanitrate (SHN)
Xylitol		Xylitol pentanitrate (XPN)

[a] Reportable explosive precursor; [b] Also known as hexamine; [c] Regulated ≥ 30 % w/w

1.1.2.2 Degradation/transformation products of explosives

Some of the precursors/intermediates shown in Table 1.3 can also be formed via degradation and/or transformation of explosives. For example, the MNGs and DNGs can be formed by hydrolysis of nitroglycerin [20]. Nitrate esters are particularly susceptible to chemical aging and degradation due to the relatively weak O-NO_2 bond [20]. Therefore, identification of degradation

products may be of interest and might help match explosives synthesised at the same time, although degradation would also be affected by storage conditions and environmental factors.

Explosives have also been shown to undergo bacterial degradation/transformation in soil samples [21]. The main transformation products of TNT are 2-amino-4,6-dinitrotoluene (2-A-4,6-DNT), 4-amino-2,6-dinitrotoluene (4-A-2,6-DNT), 2,4-diamino-6-nitrotoluene (2,4-DA-6-NT) and 2,6-diamino-4-nitrotoluene (2,6-DA-4-NT) [21]. Generally, these transformation products are of more interest for environmental analysis due to their toxicity. However, degradation/transformation products may also be encountered by forensic scientists, especially when analysing post-blast soil samples [22].

1.1.2.3 Stabilisers and plasticisers

Often commercial and military explosives are combined with stabilisers and/or plasticisers. Stabilisers are added to slow decomposition and increase shelf-life, especially of nitrate ester based explosives which are inherently chemically instable [23]. Plasticisers are added to some explosives, along with polymers, to make them safer to handle, store and transport. Energetic plasticisers have increased in popularity due to improved performance [24]. Some examples of stabilisers and plasticisers are shown in Table 1.4. Identification of these components may be of value in forensic science, especially for matching explosives found at different crime scenes or attempting to identify the manufacturer.

Table 1.4: Examples of explosive stabilisers and plasticisers

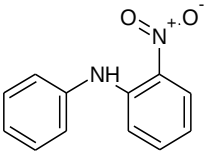
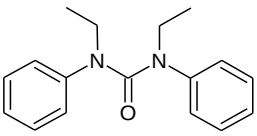
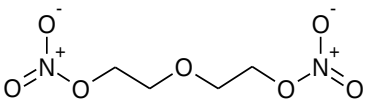
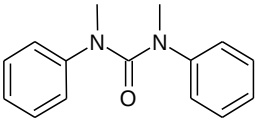
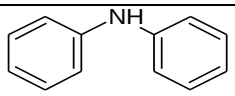
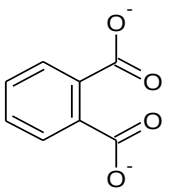
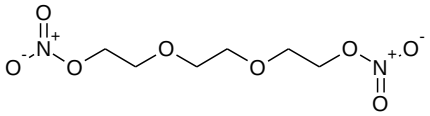
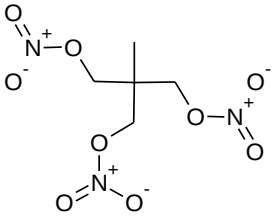
Name	Structure	Use
2-nitrodiphenylamine (2-NDPA)		Stabiliser
Diethyl diphenylurea (DEDPU)*		Stabiliser
Diethylene glycol dinitrate (DEGDN)		Energetic plasticiser
Dimethyl diphenylurea (DMDPU)**		Stabiliser

Table 1.4 (Continued): Examples of explosive stabilisers and plasticisers

Name	Structure	Use
Diphenylamine (DPA)		Stabiliser
Phthalate		Plasticiser
Triethylene glycol dinitrate (TEGDN)		Energetic plasticiser
Trimethylolethane trinitrate (TMETN)		Energetic plasticiser

*Also known as ethyl centralite, centralite or centralite I.
**Also known as methyl centralite or centralite II,

1.1.3 Requirements for forensic analysis of explosives

Since forensic scientists need to be able to identify any explosive substance, to support Explosives regulations, analytical methods are required to identify all known explosives, plus any new explosives synthesized. For forensic evidence to be used in court, the Technical/Scientific Working Group for Fire and Explosions Analysis (T/SWGFEX) recommends the “use of multiple techniques based on different principles and methodologies” for the analysis of post-blast and intact explosives [25,26]. T/SWGFEX also categorised analytical techniques, identified in a 1999 TWGFEX survey and so somewhat out of date, based on the amount of information provided. Category 1 techniques which “provide significant structural and/or elemental information”, such as liquid chromatography-mass spectrometry (LC-MS), offer the greatest confidence of identification [25,26]. More recently, the European Network of Forensic Science Institutes (ENFSI), highlighted the importance of estimating uncertainty for the forensic recovery, identification and analysis of explosives traces, which in the case of qualitative methods means assessing selectivity [27]. ENFSI also noted the current lack of consensus regarding the expression of uncertainty when identifying a chemical compound, other than to make a distinction between screening (higher uncertainty) and confirmatory (lower uncertainty) methods [27]. While confirmatory methods have a “lower degree of uncertainty”, some uncertainty remains which it is

important to estimate through appropriate selectivity experiments. The task of identifying any explosive substance is made even more challenging by the fact that traces of explosives may be all that is found on a suspect and very little explosive residue (sub-microgram [27]) remains after an explosion. Therefore, analytical methods need to have good sensitivity, as well as good selectivity. Given the number of explosive substances that may need to be identified, it would not be feasible to sequentially screen for all explosive compounds using tailored analytical methods, due to time and cost restraints. Instead, it would be desirable to have a single comprehensive screening method that could be used to detect all organic, explosive substances and direct confirmatory analysis. Ideally, a comprehensive screening method would also be able to detect precursors, intermediates, degradation/transformation products, stabilizers and plasticizers which may also be of value to forensic scientists.

1.2 Screening and identification of explosives using high performance liquid chromatography coupled to mass spectrometry (HPLC-MS)

A wide range of methods are used for explosives detection from canine olfaction to laboratory-based analytical technology. Chromatography coupled to mass spectrometry is arguably the most powerful technique for confirmatory identification of explosives due to the combined power of chromatographic separation along with the structural information provided by mass spectrometry (MS). Chromatography has the potential to separate analytes, both from other compounds that are indistinguishable by the detector, such as ions with the same mass, and from matrix components which may interfere with detection, for example by ion suppression in mass spectrometry. The importance of chromatographic separation and/or high resolution mass spectrometry for confirmatory identification of explosives was highlighted by Sisco et al. who used direct analysis in real-time (DART) MS for rapid explosives screening [28]. Confirmatory identification of specific explosive compounds was not always possible due to similar fragmentation, isobaric ions with the same nominal mass but different exact mass and isomers with identical exact mass. In the case of isobaric ions, the use of high resolution mass spectrometry would allow isobaric ions to be distinguished and in all cases combining with a chromatographic separation could enable confirmation.

Gas chromatography (GC) has been used successfully for explosives detection [29-31]. However, the thermal instability of compounds such as HMX and PETN limits analysis time and hence the potential to separate the large numbers of compounds required in screening methods. As a result of this, high-performance liquid chromatography-mass spectrometry (HPLC-MS) is favoured and will be the focus of this thesis.

1.2.1 High performance liquid chromatography

High performance liquid chromatography (HPLC) separates compounds based on their relative affinity for a stationary phase within a column and a liquid mobile phase pumped through the column under high pressure. Compounds with a higher affinity for the stationary phase will be retained for longer, resulting in different retention times. After passing through the column, the mobile phase enters a detector. Detector response versus time is plotted to give a chromatogram, with peaks observed when analytes elute from the column. Ideally these peaks will be completely

separated allowing unambiguous analyte identification and accurate quantification using peak area. Resolution (R_s) is used to measure the degree of separation, with R_s>1.5 indicating baseline resolution between two Gaussian peaks. The equation for resolution is given in Equation 1.1.

Equation 1.1: Resolution

$$R_s = \frac{2(t_{R,2} - t_{R,1})}{(W_1 + W_2)}$$

Equation 1.2: Efficiency

$$N = \frac{L}{H}$$

Equation 1.3: van Deemter Equation

$$H = A + \frac{B}{u_x} + Cu_x$$

Resolution can be improved by increasing the difference in retention time (t_R) of two analytes, in other words the selectivity of the method, or by reducing peak width (W) through improving efficiency and reducing band broadening. Efficiency (N) is dependent on plate height (H) and column length (L), as shown in Equation 1.2. Plate height is directly proportional to the size of the solute band as it travels through the column. Hence, a smaller plate height leads to narrower peaks and greater efficiency. The factors affecting plate height are described in the van Deemter Equation (Equation 1.3) and the van Deemter Plot (Figure 1.2). The effect of multiple paths, of different lengths, through the column is described by the constant A; longitudinal diffusion by the constant B divided by the linear velocity (u_x) and finally mass transfer in and between the stationary and mobile phase is described by the constant C multiplied by the linear velocity. Therefore, low flow rates reduce efficiency due to increased longitudinal diffusion, but high flow rates also reduce efficiency due to the increased effect of mass transfer.

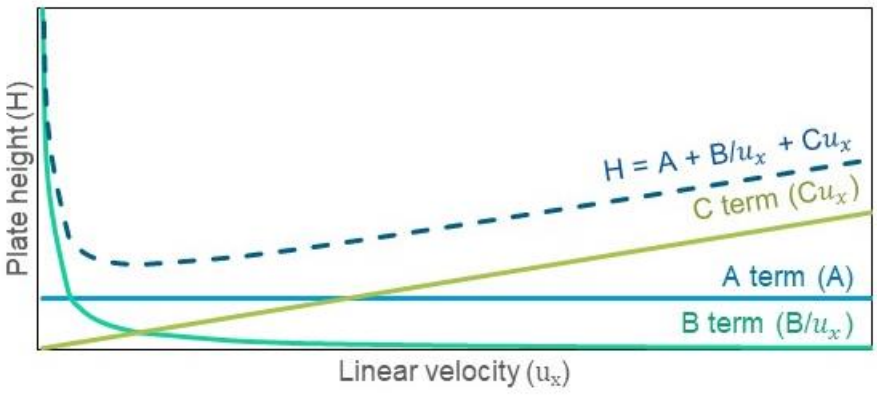


Figure 1.2: van Deemter Plot showing the effect of the A, B and C terms of the van Deemter Equation on plate height with increasing linear velocity.

As shown in a 2007 review by Gaurav et al.[56], a large number of studies have used HPLC for explosives analysis. However, the number of analytes included in these methods was relatively small, especially in comparison to the number of known explosives. Additionally, baseline separation of forensically-relevant analytes is still lacking. LC methods used for studies analysing multiple organic explosives are described in Table 1.5. Many of these methods were developed for environmental analysis and so focussed on separation of the compounds listed in the United States Environmental Protection Agency (US EPA) Method 8330, since updated to methods 8330A and B. Method 8330 and 8330A do not contain any nitrate esters [57,58]. Instead they also include TNT transformation products, 2-amino-4,6-dinitrotoluene (2-A-4,6-DNT), 4-amino-2,6-dinitrotoluene (4-A-2,6-DNT), 1,3,5-trinitrobenzene and 1,3-dinitrobenzene. While method 8330B does include two nitrate esters, many forensically relevant explosives, such as ETN and EGDN, are still missing [59].

The most common stationary phase used for organic explosives analysis is reversed-phase octadecylsilica (C_{18}). However, C_8 [49], bridged ethyl hybrid C_{18} [39,43], porous graphitic carbon [36-38], amide [53], phenyl [45] and diphenyl [50] phases have also been used in attempts to improve selectivity. Marple and LaCourse combined two different C_{18} columns in series in order to achieve separation of dinitrotoluene isomers 2,4-and 2,6-DNT and 2-A-4,6-DNT, since different pairs co-eluted on each column [55]. The combined columns also result in an increased length of stationary phase and hence increased efficiency (Equation 2). Columns with smaller particle sizes, such as 1.7 µm, have been used by both Thomas et al. and Oehrle [39,43] to improve efficiency by reducing the effect of multiple paths and mass transfer. These are classed as ultra-high performance liquid chromatography (UHPLC) columns as their particles are smaller than 2 µm and they require an HPLC system compatible with ultra-high pressure due to increased backpressure with decreased particle size. Paull et al. used monolithic columns, with lower backpressure than the more common particle packed columns, in order to use higher flow rates and achieved baseline separation of seven explosives with a total run time of 3 minutes [52]. Apart from one study using photo-assisted electrochemical detection (PAED) [55], the detectors used for the methods outlined in Table 1.5 were ultra-violet light (UV) and/or mass spectrometry (MS). The most common wavelengths used for UV detection were 254 and 210 nm. As nitrate esters do not absorb at higher wavelengths they are only detected at 210 nm, but since this wavelength is less specific than 254 nm there is a greater chance of interference from matrix

Table 1.5: Liquid chromatography methods for the separation of organic explosives.

Reference	Analytes	Eluent	Column	Temp. (°C)	Flow rate (mL min^{-1})	Total run time (min)	Inj. Vol. (µL)	LOD* (ng)	Detector	Comments
DeTata 2013 [32]	24 nitroaromatics, 9 nitrate esters, 4 nitramines, 4 peroxides, 20 others	Multistep gradient - CH_3OH:H_2O, 10 mM NH_4CHO_2	Phenomenex Kinetex C_{18} (4.6 x 150 mm, 2.6 µm)	40	1	13	10	0.1-1000	HRMS	Separation of the 17 USEPA Method 8330B analytes in less than 6 min Baseline separation only attempted for isomers
Xu 2014 [33]	15 nitroaromatics, 4 nitrate esters, 3 nitramines	isocratic - 45:55 H_2O: CH_3OH, washed with 100% CH_3OH	LichroSpher 100, RP18 (2 x 250 mm, 5 µm) and Phenomenex Securityguard C_{18} (2 x 4 mm)	35	0.2	35	10	0.011- 2.6	UV and HRMS	2,4/2,6-DNT not baseline resolved EGDN/DEGDN, TNT/3,4-DNT and tetryl/Cl-20/1,2DNB/1,4DNB do not appear fully resolved, but unclear from stacked EICs.
	TATP and HMTD	Multistep gradient - CH_3OH:H_2O, 2.5 mM NH_4OAc	GraceSmart RP18 (2.1 x 150 mm, 3 µm) and guard column	35	0.2	24	10	0.049-0.13	UV and HRMS	Separate method for peroxides
Zhao 2002 [34]	9 nitrate esters	Isocratic- 70:30 CH_3OH:H_2O	Restek Allure C_{18} (2.1 x 100 mm, 5 µm)	n.s.	0.15	7	10	0.05 - 100	MS	Nitrate esters only
		Isocratic- 70:30 CH_3OH:H_2O	Restek Allure C_{18} (3.2 x 150 mm)	n.s.	0.4	7	10	0.1 - 25		
Mathis 2005 [35]	1 nitroaromatics, 3 nitrate esters, 2 nitramines	Isocratic- 50:50 CH_3OH:H_2O	Agilent Hypersil C_{18} (2.1 x 100 mm)	n.s.	0.15	10	10	n.s. < 1-600	MS	NG and TNT not separated Poor retention of EGDN (k=0.5)
Tachon 2007 [36]	12 nitroaromatics, 2 nitrate esters, 2 nitramines	Multistep gradient – 1 mM NH_4CHO_2, H_2O: CH_3CN: IPA	PGC Hypercarb (2.1 x 100 mm, 5 µm)	70	0.2	60	10	0.4-10.6	MS	RDX/NG and 2,6/2,4 DNT not baseline resolved.
Tachon 2008 [37]	6 nitroaromatics, 3 nitrate esters, 2 nitramines	Isocratic- 55:45 CH_3OH:H_2O	C_{18} Allure (4.6 x 250 mm, 5 µm)	25	1	30	20	n.s. < 20	UV (λ=210 nm)	Separation of 2,4- and 2,6-DNT due to column retentivity rather than selectivity.
		Multistep gradient – 1 mM NH_4CHO_2, H_2O:CH_3CN:IPA	PGC Hypercarb (2.1 x 100 mm, 5 µm)	60	0.2	60	10	n.s. < 10	UV and MS	Baseline separation achieved with high carbon load (27%) C_{18} allure

n.s. not specified, n.d. not detected

*For comparison purposes, where limits of detection were not specified LOQs are given or they are estimated to be less than the amount detected in the lowest standard.

Table 1.5 (Continued): Liquid chromatography methods for the separation of organic explosives.

Reference	Analytes	Eluent	Column	Temp. (°C)	Flow rate (mL min^{-1})	Total run time (min)	Inj. Vol. (µL)	LOD* (ng)	Detector	Comments
Holmgren 2005 [38]	7 nitroaromatics, 1 nitrate ester, 13 others	Multistep gradient – H_2O:CH_3OH: CH_3CN: CH_2Cl_2:toluene.	Hypercarb PGC (4.6 x 250 mm, 5 µm)	30	0.9 - 1.4	50	n.s.	0.5- 41.2	MS	Report complete separation, but no chromatogram showing baseline resolution for all. With faster separation (17min), 3,4-DNT, 2,3-DNT, 1,3-DNB and 1,4 DNB were not resolved but distinguished by MS.
Thonas 2013 [39]	7 nitroaromatics, 1 nitrate ester, 13 others	Multistep gradient – H_2O: CH_3CN:CH_3OH, 2 mM NH_4OAc + 0.2 mM NH_4Cl	bridged ethyl hybrid (BEH) C_{18} (2.1 x 100 mm, 1.7 µm). Plus BEH C8 guard column	40	0.5	n.s. (>8)	20	0.4-64	MS/MS	3,4-DNT, 2,3-DNT, and 2-NT co-eluted as one peak NG and 4-NT not resolved.
Xu 2004 [40]	14 nitroaromatics, 4 nitrate esters, 3 nitramines	Isocratic - 50:50 CH_3OH:H_2O, with/without 2.5 mM NH_4OAc or 1 mM glycine	Waters Nova-Pack C_{18} (3.9 × 150 mm, 4 µm)	n.s.	0.4	24	10	0.012-16	MS	Identification by MS not LC separation. 2,4-DNT/2,6-DNT unresolved.
Song 2009 [41]	8 nitroaromatics, 2 nitramines	Isocratic - 50:50 CH_3OH:H_2O, with 1 % CH_2Cl_2 for nitramines	218TP52 C_{18} (2.1 x 25 mm)	n.s.	0.2	10	5 µL loop	n.s. (LOQs 0.029-0.361)	HRMS	Nitroaromatics and nitramines analysed using different methods Baseline separation of isomers but tetryl/1,3-DNB, and 2-A-4,6-DNT/2,6-DNT not fully resolved.
Bečanová 2010 [42]	14 nitroaromatics	Isocratic - 43:57 CH_3OH:H_2O with 5 mM NH_4OAc	Acclaim Explosives E1 (4.6 × 250 mm 5 µm)	25	1	70	1	0.003 -0.058	MS/MS	Explosives E1 column designed for separation USEPA Method 8330 analytes Baseline separation
Oehrle 2008 [43]	12 nitroaromatics, 2 nitrate esters, 3 nitramines 1 peroxide	Isocratic - 28:72 CH_3OH:H_2O with NH_4OAc	Acquity UPLC BEH C_{18} (2.1 x 100 mm, 1.7 µm)	55	n.s.	10.5	n.s.	low pg	UV and MS	Baseline separation (excluding nitrate esters) with UV, not MS Gradient used with nitrate esters, NG and tetryl not baseline resolved.
Perret 2008 [44]	TNT, RDX, PETN, NG, DPA and EC	Isocratic - 70:30 CH_3CN:H_2O with 1 mM NH_4OAc	X Terra C_{18} (2.1 x 150 mm, 3.5 µm)	n.s.	0.2	12	5 µL loop	0.005-4	MS/MS	Very limited separation, identification based on MS/MS

n.s. not specified, n.d. not detected

*For comparison purposes, where limits of detection were not specified LOQs are given or they are estimated to be less than the amount detected in the lowest standard.

Table 1.5 (Continued): Liquid chromatography methods for the separation of organic explosives.

Reference	Analytes	Eluent	Column	Temp. (°C)	Flow rate (mL min^{-1})	Total run time (min)	Inj. Vol. (µL)	LOD* (ng)	Detector	Comments
Ochsenbein 2008 [45]	5 nitroaromatics, 2 nitrate esters, 2 nitramines	Multistep gradient – H_2O:CH_3OH 2.5 mM NH_4OAc	Xbridge phenyl (2.1 x 150 mm, 3.5 µm)	40	0.2	23	100	n.s. LOQs 3-100	MS	No chromatogram shown but resolution factors of 1.1 for the DA-NTs and 0.9 for A-DNTs. Resolution of non-isomeric compounds not discussed.
Levsen 1993 [46]	15 nitroaromatics, 1 nitrate esters, 2 nitramines, 1 other	Isocratic - 57:43 CH_3OH:H_2O	Spherisorb C_{18} (2.4 X 250 mm, 5 µm)	n.s.	0.4	40	30	0.8-4.6	UV (λ=210 or 254 nm)	PA/Hexyl, 2-A-6-NT/1,3,5-TNB and 2,4,6-TNT/4-A-2,6-DNT not baseline separated.
DeTata 2013 [47]	13 nitroaromatics, 2 nitrate esters, 2 nitramines	isocratic - 43:57 CH_3OH:H_2O	Acclaim E1 column (4.6 x 250 mm, 5 µm)	32	1	n.s.	n.s.	n.s.	UV (λ=210 and 254 nm)	Co-elution of NG and tetryl.
Liu 2007 [48]	12 nitroaromatics, 2 nitramines,	Isocratic - 43:57 CH_3OH:H_2O	Acclaim Explosives E1 (4.6 × 250 mm, 5 µm)	30	1	40	5	0.2-4.6	UV (λ=254 nm)	Baseline separation of the 14 anayltes in method 8330 Dasing an isocratic method.
	13 nitroaromatics, 2 nitrate esters, 2 nitramines, 1 other	isocratic - 48:52 CH_3OH:H_2O	Acclaim Explosives E2 column (4.6 × 250 mm5 µm)	30	1	50	5	0.8-4.1	UV (λ=210 nm)	Resolution > 1.8 for all
Borch 2004 [49]	16 nitroaromatics, 2 nitramines, 5 others	Multistep gradient – H_2O:CH_3OH, 0.025 M sodium phosphate	Supelcosil octyl C-8 (4.6 x 150 mm, 5 µm) Supelcosil LC-8 (4.6 x 20 mm, 5 µm)	50.5	1	34	10	n.s.	UV (λ=220, 230, 254, 360, and 370 nm)	Baseline separation of EPA method 8330 compounds When TNT metabolites were added, separation no longer baseline.
Lordel-Madeleine 2013 [50]	6 nitroaromatics	Isocratic – 62:38 CH_3OH:H_2O	Symmetry Shield RP18 (3 x 250 mm, 5 µm)	n.s.	0.4	10	10	LOQ 200–550	UV (λ=228, 237 and 247 nm)	Resolution of 2,4- and 2,6-DNT was 1.25 on SymmetryShield RP18 column, 1.18 on the JSphere C18 column, and 1.64 on the diphenyl column.
		Isocratic – 62:38 CH_3OH:H_2O	JSphere H80 (3 x 150, 4 µm)	n.s.	0.4	15	10	LOQ 160–420 ng/µL		
		Multistep gradient – H_2O:CH_3CN	diphenyl XRS (3 x 100 mm, 5 µm)	n.s.	0.4	25	10	LOQ 360–1000		

n.s. not specified, n.d. not detected

*For comparison purposes, where limits of detection were not specified LOQs are given or they are estimated to be less than the amount detected in the lowest standard.

Table 1.5 (Continued): Liquid chromatography methods for the separation of organic explosives.

Reference	Analytes	Eluent	Column	Temp. (°C)	Flow rate (mL min^{-1})	Total run time (min)	Inj. Vol. (µL)	LOD* (ng)	Detector	Comments
Tyrrell 2011 [51]	12 nitroaromatics, 2 nitrate esters, 2 nitramines	isocratic – 48:52 $CH_3OH:H_2O$	Acclaim Explosives E2 (3 x 150 mm, 3 µm)	30	0.42	30	5 µL loop	0.6-6.9	UV (λ=210 nm)	TNT/tetryl, 4-A-2,6-DNT/2-A-4,6-DNT and 4-NT/3-NT not baseline Minimum resolution was 0.92 for the 3-NT and 4-NT peak pair
	4 nitroaromatics, 3 nitrate esters, 2 nitramines	Multistep gradient – $CH_3OH:H_2O$	Acclaim Explosives E2 (3 x 150 mm, 3 µm)	32	0.7	23	5 µL loop	1.55-7.7	UV (λ=210 nm)	Minimum resolution was 1.16 for tetryl/NG, and NG/TNT also only 1.23. 2,4- and 2,6-DNT co-eluted
Paull 2004 [52]	9 nitroaromatics, 1 nitrate esters, 2 nitramines	Linear gradient - $CH_3OH:H_2O$	Chromolith SpeedROD C_{18} (4.6 x 50 mm)	room temp.	3	n.s. (> 6)	10	0.21-0.94	UV (λ=210 and 254 nm)	2,3- 2,4- and 2,6-DNT coelute.
		Linear gradient - $CH_3OH:H_2O$	Chromolith SpeedROD C_{18} (4.6 x 50 mm) Chromolith Performance C_{18} (4.6 x 100 mm)	60	3	n.s. (> 14)	25	0.25-1.04	UV (λ=210 and 254 nm)	Baseline separation without 2,4-DNT
		Linear gradient - $CH_3OH:H_2O$	Chromolith SpeedROD C_{18} (4.6 x 50 mm)	60	8	3	n.s.	n.s	UV (λ=210 and 254 nm)	Baseline separation of HMX, RDX, TNT, 2,4-DNT, 2-NT, 3-NT and PETN Peak widths only 7-10 s
Gaurav 2009 [53]	11 nitroaromatics	isocratic - 43:57 $CH_3OH:H_2O$	Supelco Ascentis RP amide column (4.6 x 150 mm, 5 µm)	n.s.	1.3	25	n.s.	0.17-0.93 pg/µL	UV (λ=254 nm)	Baseline resolution but broad peaks
Reifenrath 2002 [54]	9 nitroaromatics, 2 nitramines 1 other	Isocratic – 50:50 $CH_3OH:H_2O$	Supelcosil LC-18-S (4.6 x 250 mm, 5 µm), Supelguard LC-18-S (20 mm)	n.s.	1	15	50	n.s.	UV (λ= 254 nm) and radio-metric flow detector	Large retention time bracket for dinitroaniline (10-12 min) over laps with TNT (12.00 min) and tetryl (10.09 min). 4-A-2,6-DNT, 2-A-4,6-DNT, 2,6-DNT and 2,4-DNT have close retention times (12.55, 13.14, 13.31 and 13.47 min) but no peak widths or chromatograms.

n.s. not specified, n.d. not detected

*For comparison purposes, where limits of detection were not specified LOQs are given or they are estimated to be less than the amount detected in the lowest standard.

Table 1.5 (Continued): Liquid chromatography methods for the separation of organic explosives.

Reference	Analytes	Eluent	Column	Temp. (°C)	Flow rate (mL min^{-1})	Total run time (min)	Inj. Vol. (µL)	LOD* (ng)	Detector	Comments
Marple 2005 [55]	13 nitroaromatics, 3 nitrate esters, 2 nitramines, 1 other	Isocratic – 50:50 CH_3OH:H_2O with 20 mM NH_4OAc	Denali (4.6 × 250 mm, 5 µm) and SelectaPore (4.6 × 250 mm, 5 µm)	30	1	35	100 µL loop	PAED 0.007−3 pg/µL UV 0.9−5 pg/µL	UV and PAED	2,6-DNT/2-A-4,6-DNT coelute on one column, 2-A-4,6-DNT/2,4DNT on the other Coupling 2 columns improved separation, not quite baseline for 4-A-2,6-DNT/2,6-DNT/2-A-4,6-DNT/2,4DNT (nitro esters and PA not included) On single column NQ unretained, and EGDN/RDX and 2,6-DNT/2-A-4,6-DNT not fully resolved.

n.s. not specified, n.d. not detected
*For comparison purposes, where limits of detection were not specified LOQs are given or they are estimated to be less than the amount detected in the lowest standard.

components. Tachon et al. made use of the reduced selectivity at 210 nm in order to evaluate the effectiveness of their solid phase extraction (SPE) procedure for the removal of interfering compounds from a motor oil matrix that coelute with explosive compounds and may affect detection [37].

Since MS can distinguish between ions with different *m/z*, many LC-MS methods only attempt separation of isomers which cannot be distinguished by MS. However, for forensic purposes the extra confidence of identification achieved from both chromatographic separation and mass spectra is desirable, particularly when several explosives form common ions.

1.2.2 Mass spectrometry

Mass spectrometers separate and detect ions based on their mass to charge ratio (*m/z*). There are three main components of a mass spectrometer, the ionisation source where gas-phase ions are produced, the mass analyser where these ions are separated based on their *m/z* and the detector. Ideally all ions would be transferred from the ion source to the mass analyser, but in reality, a large percentage may be lost along the way. Various ion guides, lenses and funnels have been designed to minimise these losses and hence improve sensitivity.

1.2.2.1 Ionisation

While mass spectrometry is often considered a universal detector, only charged molecules can be detected. Effective ionisation is crucial for an analyte to be detected. Atmospheric pressure ionisation sources are often used when coupled to HPLC, as can be seen in Table 1.6. While Song and Bartmess used atmospheric pressure photoionization (APPI), with UV radiation promoting ionization [41]; electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI) are the most common ionisation sources used for the detection of explosives.

A schematic of ESI is shown in Figure 1.3. In ESI, the first stage in formation of gas phase ions is the production of charged droplets at the tip of the electrospray needle [60]. A solution is passed through a spray needle into an electric field which leads to partial separation between positive and negative ions in the solution. In negative ionisation mode, a high negative potential is applied to the tip of the needle and an electrochemical reduction reaction creates negative ions or removes positive ions at the needle tip. Repulsion of the negative ions at the tip of the spray needle and attraction towards the counter electrode results in formation of a Taylor cone causing the charged droplets to leave the needle tip. These droplets reduce in size until the Rayleigh limit

Table 1.6: Detection details for LC-MS analysis of multiple explosives

Reference	Analytes*	LOD (ng on column)	Ionisation	Mass analyser	Additives (pre/post column)	Main positive ions	Main negative ions	MS/MS used	Mass resolution†
DeTata 2013 [32]	24 nitro-aromatics, 9 nitrate esters, 4 nitramines, 4 peroxides, 20 others	0.1-1000	APCI (+ and –)	QToF	10 mM ammonium formate (pre)	$[M+H]^+$, $[M+NH_4]^+$	$[M+CHO_2]^-$, $[M-H]^-$ $[M]^-$, $[NO_3]^-$, $[M]^-$ $NO_2+CH_2O_2]^-$, $[M+NO_3]^-$	yes	High (∼20,000 resolving power)
Xu 2014 [33]	15 nitro-aromatics, 4 nitrate esters, 3 nitramines	0.011-2.6	APCI (-)	LTQ-Orbitrap	4% v/v chloroform in methanol (post)	n/a	$[M+Cl]^-$, $[M-H]^-$, $[M]^-$	yes	High (> 60,000 resolving power)
	2 peroxides	0.049-0.13	APCI (+)	LTQ-Orbitrap	2.5 mM ammonium acetate (pre)	$[M+H]^+$, $[M+NH_4]^+$	n/a		
Zhao 2002 [34]	9 nitrate esters	NH4NO3 0.05-20 NaNO2 0.1-100 CH3CH2COOH 0.05 -n.d. NH4Cl 0.1-25	ESI (-)	LCQDUO ion trap	0.05 mM ammonium nitrate, 0.1 mM sodium nitrite, 0.2 mM propionic acid or 0.1 mM ammonium chloride (post)	n/a	$[M+NO_3]^-$, $[M+NO_2]^-$ $[M+CH_3CH_2CO_2]^-$ or $[M+Cl]^-$	yes	Low
		0.1 - 25	APCI (-)	LCQDUO ion trap	0.2 % (v/v) dichloromethane, 0.1 % (v/v) chloroform, 0.05 % (v/v) carbon tetrachloride or 0.3 mM ammonium chloride (post)	n/a	$[M+Cl]^-$		
Mathis 2005 [35]	1 nitro-aromatics, 3 nitrate esters, 2 nitramines	n.d. 10 µg/mL standards (EGDN 60 µg/mL)	ESI (-)	Quad-rupole ion-trap	0.05 mM ammonium nitrate and 0.1 mM ammonium chloride (post). OR 0.05 mM nitrate, 0.1 mM ammonium chloride, 0.1 mM ammonium formate and 0.1 mM ammonium acetate (pre)	n/a	$[M+NO_3]^-$, $[M+Cl]^-$ $[M+CHO_2]^-$, $[M+OAc]^-$	no	Low
Tachon 2008 [37]	6 nitro-aromatics, 3 nitrate esters, 2 nitramines	n.d. 1 ng/µL standard used	APCI (-)	Ion trap LCQ Ad-vantage	1 mM ammonium formate (pre)	n/a	$[M+CHO_2]^-$, $[M]^-$	no	Low

Table 1.6 (Continued): Detection details for LC-MS analysis of multiple explosives

Reference	Analytes*	LOD (ng on column)	Ionisation	Mass analyser	Additives (pre/post column)	Main positive ions	Main negative ions	MS/MS used	Mass resolution†
Holgren 2005 [38]	7 nitro-aromatics, 1 nitrate ester, 13 others	0.5 - 41.2	APCI (+ and -)	Bruker Esquire 3000+ ion trap	1 % (v/v) dichloromethane (pre)	$[M]^+$	$[M+Cl]^-$, $[M]^-$	yes	Low
Tachon 2007 [36]	12 nitro-aromatics, 2 nitrate esters, 2 nitramines	0.4-10.6 (using APCI)	APCI and ESI (-)	ion trap LCQ Ad-vantage	1 mM ammonium formate (pre)	n/a	$[M+CH_2O_2-H]^-$, $[M+2CH_2O_2-H]^-$, $[M]^-$ and $[M-H]^-$	no	Low
Thomas 2013 [39]	7 nitro-aromatics, 1 nitrate ester, 13 others	0.4-64	ESI (+/ -) and APCI (-) switching	Quattro Micro APITM QqQ	0.2 mM ammonium chloride and 2 mM ammonium acetate (pre)	$[M+H]^+$	$[M+Cl]^-$, $[M-H]^-$	yes	Low
Xu 2004 [40]	14 nitro-aromatics, 4 nitrate esters, 3 nitramines	0.012-n.d. (APCI-) 0.06-n.d. (ESI-) 3.1-16 (ESI+ with glycine, NTs only)	APCI or ESI mode	Finnigan MAT TSQ (QqQ)	none, 2.5 mM ammonium acetate or 1 mM glycine (pre)	$[M+glycine]^+$	$[M+OAc]^-$, $[M-H]^-$, $[M]^-$	no	Low
Song 2010 [41]	8 nitro-aromatics, 2 nitramines	n.d. LOQs 0.029-0.361 ng	APPI (-)	QSTAR XL triple QTOF	toluene (post)	n/a	$[M+Cl]^-$, $[M-H]^-$, $[M]^-$	no	high (*m/z* given to 4 dp)

Table 1.6 (Continued): Detection details for LC-MS analysis of multiple explosives

Reference	Analytes*	LOD (ng on column)	Ionisation	Mass analyser	Additives (pre/post column)	Main positive ions	Main negative ions	MS/MS used	Mass resolution†
Bečanová 2010 [42]	14 nitro-aromatics	SIM 0.002.9-0.050 MRM 0.004-0.058	ESI (-)	Agilent 6410, QqQ	5 mM ammonium acetate (pre)	$[M+H]^+$	$[M-H]^-$	yes	Low
Oehrle 2008 [43]	12 nitro-aromatics, 2 nitrate esters, 3 nitramines 1 peroxide	low pg (exact values not shown)	ESI/APCI switching	SQ	ammonium acetate (pre)	$[M+H]^+$	$[M+OAc]^-$, $[M-H]^-$	no	Low
Perret 2008 [44]	TNT, RDX, PETN, NG, DPA and EC	0.001-0.8	IonSpray (+ and -)	PE Sciex API 3000 QqQ	1 mM ammonium acetate (pre)	$[M+H]^+$	$[M+OAc]^-$, $[M-H]^-$	yes	Low
Ochsenbein 2008 [45]	5 nitro-aromatics, 2 nitrate esters, 2 nitramines	not determined, LOQ using SPE 0.03-1 ng/ul in water	ESI (+ and -)	API 5000 QqQ	2.5 mM ammonium acetate (pre)	$[M+H]^+$	$[M+OAc]^-$, $[M-H]^-$	yes	Low

*While Tetryl could be classed as a nitroaromatic and a nitramine, it has only been counted as a nitroaromatic to avoid double counting; QTof – quadrupole time of flight; QqQ – triple quadrupole
†Low mass resolution refers to unit mass resolution

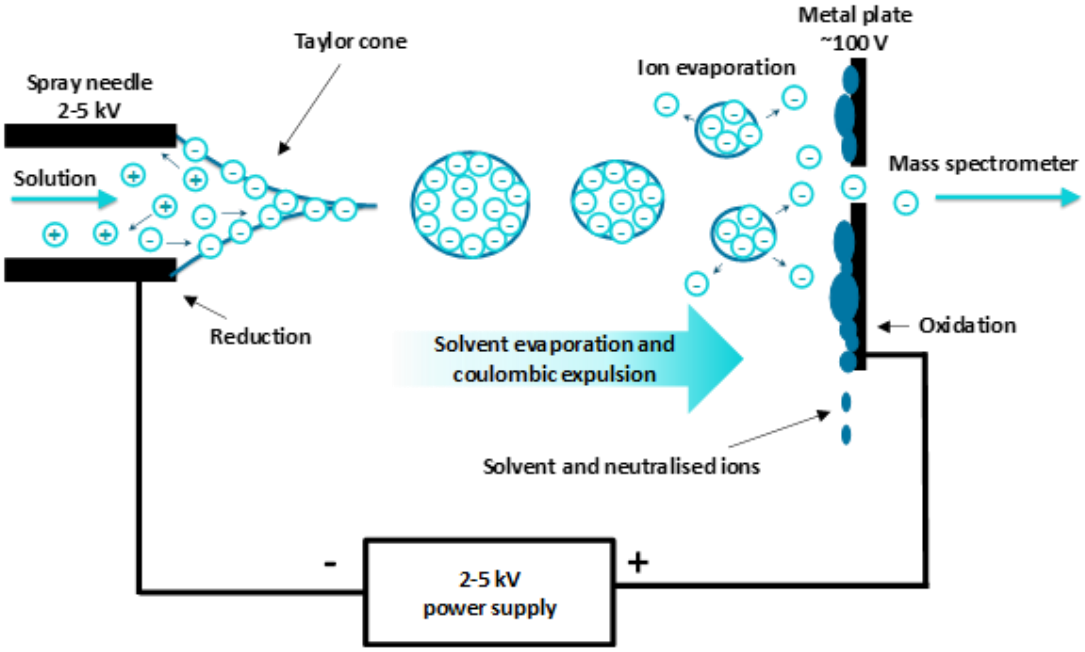


Figure 1.3: Schematic of negative electrospray ionisation, adapted from [62], [60] and [61].

is reached (Coulombic repulsion of the surface charge equals surface tension) due to solvent evaporation, resulting in Coulombic fission. These droplets reduce in size until individual gas phase ions are formed which enter the mass spectrometer. Two mechanisms have been proposed for this process, the charge residue model (CRM) and the ion evaporation method (IEM). In both CRM and IEM, droplets initially decrease in size by solvent evaporation until Coulombic repulsion overcomes surface tension, leading to expulsion of smaller charged droplets. The CRM model proposes that solvent evaporation and Coulomb fission continues until no further evaporation is possible and dry charged residues, including single ions, remain. The IEM model proposes that once the charged droplets reach a certain size, direct emission of ions into the gas phase occurs because a lower charge is required on the droplet for ion emission than for Coulomb fission. IEM is considered the more likely mechanism for the formation of small ions [60]. CRM is more likely with large ions [61]. In both models, ESI is dependent on the presence of electrolyte ions in solution. If the electrolyte ions in solution are due to the analyte, then gas-phase analyte ions will be produced. However, at high electrolyte concentrations the conversion to gas-phase ions is less efficient and the presence of buffers or other mobile phase additives in solution will compete with the analyte [60]. For the same reason ESI is susceptible to ion suppression due to matrix [61].

In APCI (Figure 1.4), the solution passes through a needle/capillary similar to that used in ESI, but with APCI the capillary is not charged. Instead, the first stage of APCI is vaporisation in a

heated chamber. Gas-phase molecules then pass a high voltage corona discharge needle which promotes ionisation of solvent molecules. The primary (solvent) ions then cause secondary ionisation of the analyte, as shown in the box in Figure 1.4: Schematic of positive atmospheric pressure chemical ionisation. Equations outlining the ionisation mechanism are from Awad et al. with S referring to the vaporised solvent and M referring to the analyte [61].Figure 1.4. In positive ion mode APCI, mechanisms for ionisation of the analyte include: charge exchange or electron transfer from a neutral analyte molecule (M) to a solvent ion ($S^{+\cdot}$); proton transfer from the solvent (S) to the analyte and adduct formation. In negative mode, mechanisms include: proton abstraction from the analyte by the solvent; associative electron capture, when the analyte gains an electron; dissociative electron capture, when the analyte gains an electron and fragments, and adduct formation [63].

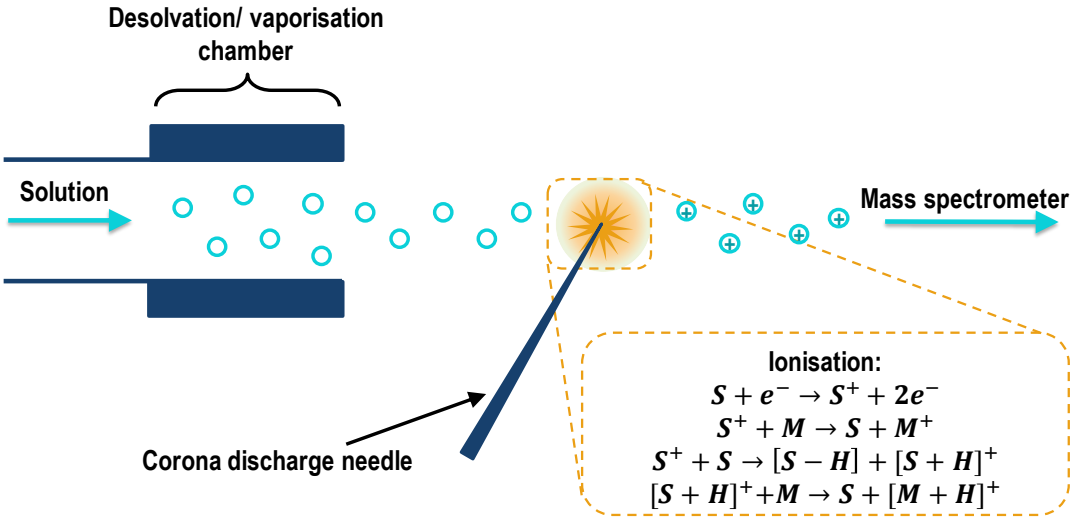


Figure 1.4: Schematic of positive atmospheric pressure chemical ionisation. Equations outlining the ionisation mechanism are from Awad et al. with S referring to the vaporised solvent and M referring to the analyte [61].

A significant difference between ESI and APCI is that in ESI the analyte is ionised in the liquid phase before being released into the gas phase, whereas in APCI the analyte is vaporised before being ionised. As a result of this, ESI ion formation tends to be more susceptible to non-volatile matrix components that can lead to ion suppression due to changes in droplet properties or competition for charge or space on the droplet surface [61]. The effect of this can be minimised by chromatographic separation from major matrix components. Additionally, the requirement for high vaporisation temperatures with APCI may be problematic for thermally unstable compounds. Another difference between the two ionisation techniques is that ESI is typically more efficient for

ionisation of polar analytes, whereas APCI tends to be better for low-to-medium polarity analytes [61].

Two of the main types of ion formed in both ESI and APCI are the protonated and deprotonated molecules, $[M+H]^+$ and $[M-H]^-$. Analytes containing amine groups, such as the explosive hexamethylene triperoxide diamine (HMTD) and stabilisers like diphenylamine (DPA), are basic and so readily form $[M+H]^+$ ions in positive mode [33,39]. In contrast, most explosives containing nitro groups ionise more readily in negative mode. Nitrate esters (e.g. NG) and nitramines (e.g. RDX) have weak deprotonation due to low gas phase acidities. For these analytes the formation of adducts such as $[M+OAc]^-$, $[M+Cl]^-$ or $[M+NO_3]^-$ is often promoted by buffers or additives, such as ammonium acetate, chloride or nitrate, to improve ionisation and hence limits of detection [34,35]. However, nitrotoluenes (e.g. TNT) generally form $[M-H]^-$ (and/or $[M]^-$ in APCI) and ionise most efficiently with unbuffered mobile phases as the presence of buffers or additives leads to ion suppression [40]. Mono-nitrotoluenes (NTs) are particularly challenging to ionise and as such are missing from many of the MS methods shown in Table 1.6. Xu et al. found the formation of a glycine adduct in positive mode, $[M+glycine]^+$, improved limits of detection and proposed confirmation by this method where an NT peak was detected by UV, which still had superior sensitivity to MS for these analytes [64]. Adducts are also used in positive mode for the detection of peroxides such as TATP and MEKP which are often detected as the ammonium adduct, $[M+NH_4]^+$ [32,33].

Since different classes of organic explosive ionise best under different conditions, several methods change ionisation parameters throughout the run by switching between positive and negative mode [32,38] or even switching between APCI and ESI [39,43]. However, switching polarity or between APCI/ESI takes additional time and so results in a slower total scan time and fewer data points per chromatographic peak. This may be problematic, particularly when coupled to UHPLC which has narrower peaks than conventional HPLC. Changing ionisation mode may also affect quantification by leading to variability in peak area, as found by Thomas et al. who reported 43 % relative standard deviation (RSD) for NG [39]. Other studies used multiple methods [33,40], allowing ionisation to be optimised for specific groups of analytes but resulting in more time-consuming analysis. Xu et al. decided against the use of dynamic polarity switching for their recent screening method as it decreased the mass accuracy of the Orbitrap Fourier transform

(FT) MS detection and instead used separate positive and negative methods which also allowed the use of different ionisation temperatures [33].

1.2.2.2 Mass analysers

Mass analysers separate ions in space or time based on their mass to charge ratio through the use of electric or magnetic fields. A range of different mass analysers have been used for the detection of explosives including time of flight (TOF), quadrupole (Q), ion trap and Orbitrap mass analysers (Table 1.6). Time of flight analysers separate ions based on the time taken for ions to reach a detector, due to the slower velocity of ions with larger *m/z*. Quadrupole mass analysers separate ions based on their oscillation between four parallel rods to which a constant voltage and radio frequency oscillating voltage are applied. In the case of transmission quadrupoles, a two-dimensional electric field is set up that pushes or pulls ions of a particular *m/z* towards the detector with ions of other *m/z* neutralised through collision with one of the rods. In the case of a quadrupole ion trap, a three-dimensional electric field is set up resulting in all *m/z* ions being trapped in the quadrupole. Ions can be separated as they leave the trap based on their *m/z*, with the lowest *m/z* ions leaving first.

Orbitrap mass analysers (Figure 1.5) trap ions between one inner and two outer electrodes using static electrostatic fields. Short packets of ions are injected into the Orbitrap with a velocity perpendicular to the z-axis. A radial electric field traps the ions radially and they orbit around the inner electrode, giving the Orbitrap its name. The ion injection point is offset from z=0 and an axial electric field causes ions to also oscillate along the z axis in a harmonic motion. The frequency of

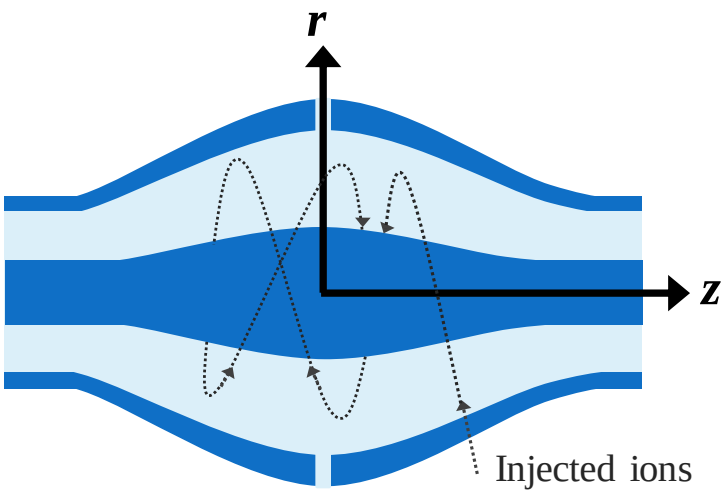


Figure 1.5: Cutaway view of an orbitrap mass analyser. Ions are injected perpendicular to the z-axis, displaced from z=0. This gives the ions potential energy in the z-direction causing them to oscillate along the z-axis. Adapted from Hu et al. 2005 [65].

oscillations along the z-axis is dependent on *m/z*, and independent of energy and the spatial spread of ions. This allows ions with differing *m/z* to be distinguished with high mass accuracy and resolution [65].The two outer electrodes are used as receiver plates for broadband image current detection of the ions and their axial oscillations. The recorded time-domain signal is then Fourier transformed into the frequency-domain and converted into a mass spectrum [66].

The International Union of Pure and Applied Chemistry (IUPAC) defined mass resolution (*m/Δm*) as the mass of the ion of interest (*m*) divided by the difference in mass (*Δm*) that can be separated [67]. The difference in mass (*Δm*) can either be determined by the difference in mass between two adjacent peaks of equal height or based on the peak width of a single peak. When two adjacent peaks are used, the two peaks must have a valley of <10 % peak height between them [68]. For the peak width definition, *Δm* commonly refers to the full peak width at half maximum peak height (FWHM) [68]. It is the FWHM mass resolution definition that will be used in this thesis. Since resolution varies based on the *m/z* being measured (*m*), with fixed resolution, smaller differences in mass (*Δm*) can be separated for smaller ions. Quadrupoles are low resolution mass analysers and commonly said to have unit mass resolution which actually refers to the *Δm* that can be separated, rather than *m/Δm.* This means the $[M-H]^-$ ion for NG ($C_3H_5N_3O_9$, MW=227.00258) and TNT ($C_7H_5N_3O_6$, MW=227.01784) would appear at *m/z* 226 for both and be indistinguishable using a quadrupole mass analyser. However, they could be distinguished using a high-resolution mass analyser such as TOF or Orbitrap.

Providing that the mass analyser also has high mass accuracy, high resolution mass spectrometry (HRMS) can distinguish between compounds with the same nominal mass and provides higher confidence of ion identification. Good mass resolution means that two ions with a small mass difference can be separated. However, good mass accuracy is also required to determine which ion is present, if only one peak is detected. When run in full scan mode, HRMS has the potential for suspect screening and non-target detection where post-acquisition data mining can be used for the detection of new target analytes as the need arises. Orbitraps are able to achieve even higher resolution than TOF mass analysers, with Orbitraps offering a resolving power of up to 140,000 at full width half maximum (FWHM) [69] compared to 40,000 for TOF mass analysers [70]. Providing mass accuracy is also high, increased resolving power reduces the number of compounds with the same mass, increasing confidence in ion identification. However, it also

comes at the cost of longer scan speeds which may limit the compatibility of Orbitrap mass analysers with UHPLC, due to narrow peak widths.

Tandem mass spectrometry (MS/MS), where precursor ions are fragmented into product ions, has also been used to increase confidence in identification of specific explosives by providing additional structural information [32,39,42,44,45]. The sensitivity of MS/MS techniques measuring specific precursor-product transitions is generally very good due to low background noise, as demonstrated by Bečanová et al. who reported limits of detection (LODs) of 2.9-50 pg μL^{-1} [42]. However, the usefulness of tandem MS in broad screening methods targeting a large number of analytes is limited by the acquisition time of each transition. Additionally, only analytes with the precursor-product transitions being monitored will be detected so it is unsuitable for non-targeted analysis. This issue might be resolved through the use of all ion fragmentation (AIF) where, instead of measuring specific transitions, full scan data is taken before and after fragmentation and product ions may be associated with precursor ions based on retention time. Alternatively, MS/MS could be used for confirmation following a positive result from a screening method.

1.2.3 Liquid chromatography - high resolution mass spectrometry (LC-HRMS)

Recently, LC-HRMS has been used for the targeted analysis of multiple classes of explosive, with 24 compounds detected by Xu et al. [33] and 62 detected by DeTata et al. [32]. Although, a slightly lower mass resolution (~20,000 FWHM) was used by DeTata et al., than by Xu et al. who used a mass resolution of 60,000-100,000 FHWM. Compared to low resolution MS, high mass resolution offers increased confidence of identification due to the ability to separate isobaric ions and improved selectivity from matrix for extracted ion chromatograms, even with full scan analysis. Full-scan HRMS analysis also offers the potential for suspect or non-target screening, which is yet to be explored for explosives analysis and will be discussed further in Section 1.3. The separation power of HPLC has also not yet been fully utilised with Xu et al. only attempting baseline separation for isomers and a number of structurally similar pairs including 2,4-DNT/2,6-DNT, EGDN/DEGDN, TNT/3,4-DNT and 1,2-DNB/1,4-DNB not resolved to baseline by DeTata et al. Ionisation also remains a challenge with nitrobenzene (NB) not detected by DeTata et al. and outside the mass range of the method validated by Xu et al.

1.3 Suspect and non-target screening

With full-scan HRMS, the analyst does not need to know in advance what they are looking for, since data for all ions (within a specified range) are collected. Krauss et al. took advantage of this and proposed three different workflows for environmental analysis of polar micropollutants such as pesticides, pharmaceuticals, and industrial chemicals [71]. The three workflows were for targeted analysis, suspects screening and non-target analysis and are summarised in Figure 1.6. With targeted analysis, reference standards were analysed at the same time as samples. This enabled matching of measured retention times and HRMS data for identification purposes and, if required, quantification using calibration standards. With suspect and non-target screening, reference materials were not analysed in the first instances and so HRMS data and predicted retention times were used to generate a list of likely suspect or non-target compounds present. Ibáñez et al. made a further distinction between biased and unbiased non-target screening for

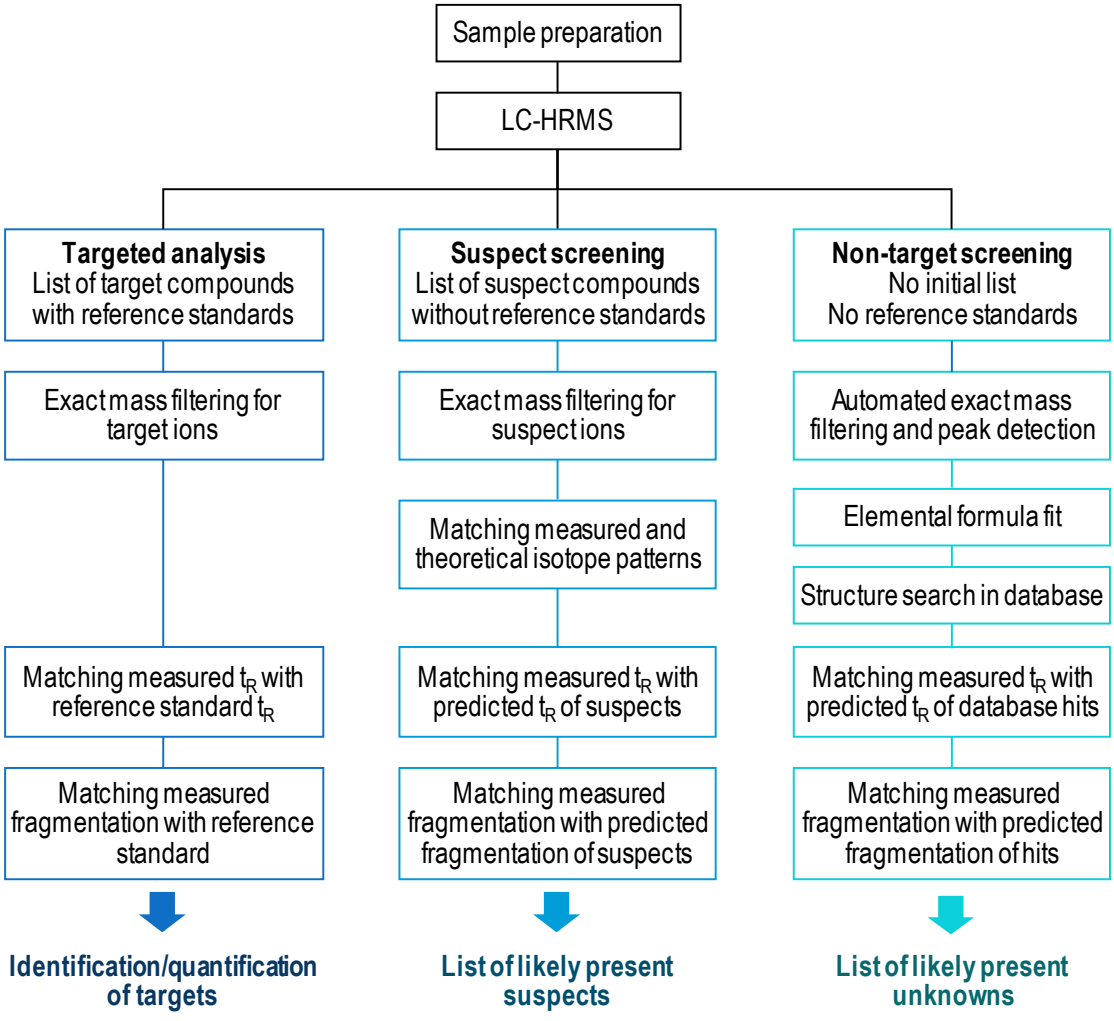


Figure 1.6: Workflows for three approached to using LC-HRMS: targeted analysis, suspect screening and non-target analysis. Adapted from ([71]).

new psychoactive substances [72]. In unbiased non-target screening, in theory the aim is to detect all components detected within a sample. In contrast, for biased non-target screening it is only structurally related unknowns that are of interest [72]. It is biased non-target screening that is of more interest for the analysis of unknown energetic materials as it is only the structurally related, energetic components of a sample that are of interest.

Development of an LC-HRMS method that could be used both for targeted analysis of commonly encountered energetic materials, suspect-screening of less commonly encountered compounds and biased non-target analysis of new compounds could be of value in forensic science. While reference materials would ultimately be required to confirm identification, especially for forensic purposes; initial screening without reference materials could enable a greater number of compounds to be detected in one method, than would be feasible if reference materials were required for all compounds. Additionally, in some cases reference materials may not be readily available. Suitable retention time prediction models may also be required for energetic materials, to reduce the number of potential positives identified by HRMS data alone.

1.3.1 Retention time prediction using quantitative structure – retention relationships

Quantitative structure – retention relationship (QSRR) prediction models use the relationship between numerical molecular descriptors and retention time to predict the retention time of either new compounds using the same chromatographic conditions or the same compounds using different chromatographic conditions [73]. The molecular descriptors may be theoretical descriptors (calculated or predicted from the chemical structure) or experimental descriptors (physical properties measured experimentally). For the prediction of new compounds, QSRR models based on theoretical descriptors are preferable as experimental descriptors may not be available and if experimental determination is required then the time and cost savings of using a QSRR model are lost and retention time may as well be obtained experimentally. The use of QSRRs for retention time prediction has been extensively reviewed by Put & Vander Heyden [73], Héberger [74], Kaliszan [75] and more recently Amos et al. [76]. Some key themes arising from these reviews are the importance of selecting appropriate descriptors, validating prediction models and determining an applicability domain, outlining descriptor values for which the model can be used without extrapolation.

Linear models have been developed for predicting reversed-phase LC retention [77-81]. Multilinear regression (MLR) models are one of the simplest ways of predicting retention time and

are easy to interpret, since the sign and size of coefficients indicates whether the descriptors increase or decrease retention time and the importance of each descriptor within the model. However, MLR models are not always able to fully describe retention and both Aalizadeh et al. and Goodarzi et al. found performance of a non-linear model to be superior [80,81]. A number of non-linear models, including artificial neural networks (ANNs) [82-84] and support vector machines (SVMs) [80,81] have also been used for revered-phase LC retention time prediction.

1.3.1.1 Artificial neural networks

Artificial neural networks (ANNs) are analogous to biological neural networks in a brain. Like biological neurons, artificial neurons are connected within a network and each neuron receives input signals (either from the original data or the output of other neurons) that affect whether or not an output signal is sent [85]. Input signals have a strength (or weight) applied to them, which is optimized during training to minimise the prediction error for the training set. A neuron will only send an output signal if the input signals exceed a threshold level, which is also optimised during training [86]. An example of a feedforward neural network is shown in Figure 1.7, where signals are transmitted from the inputs (in this case molecular descriptors) forwards through any hidden units to the output (in this case predicted retention time). Figure 1.7 is also an example of a 3-layer multilayer perceptron (MLP). In MLPs, each neuron in a particular layer is connected to each neuron in the next layer [41]. 3-layer MLPs consist of an input layer containing the independent variables, in this case molecular descriptors; one hidden layer and an output layer containing the dependent variable, in this case retention time.

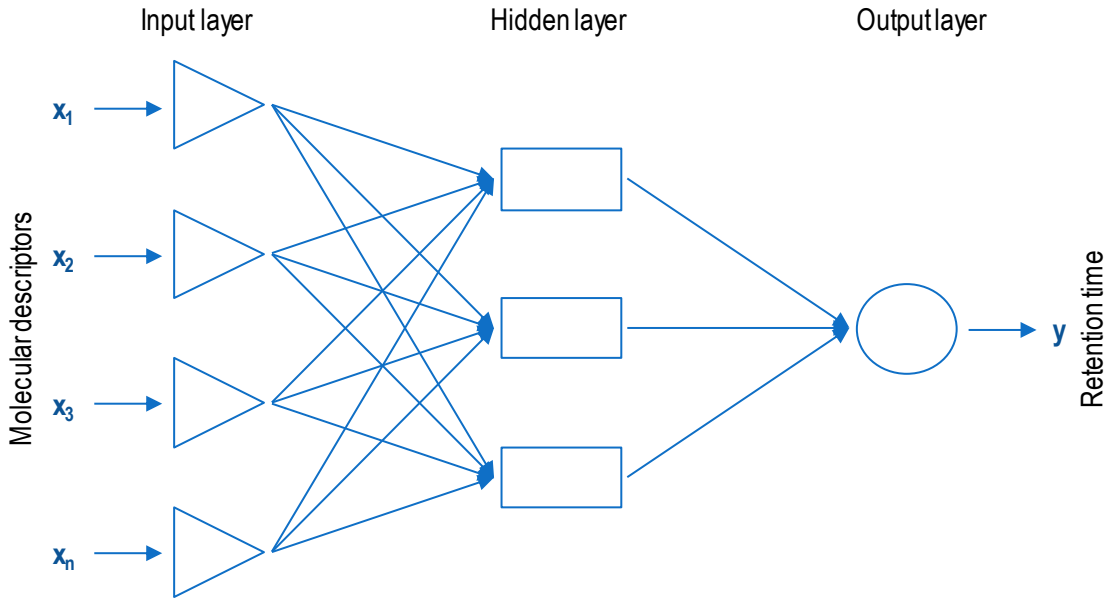


Figure 1.7: Schematic of a feed-forward ANN with an input layer, one hidden layer and an output layer containing four, three and one neurons respectively. Adapted from [73] and [87].

ANNs are one of the most well-known non-linear regression methods [87] and more flexible than multiple linear regression (MLR) since both linear and non-linear functions can be used, which may lead to improved prediction through description of more complex relationships [73]. However, the increased complexity means that ANN models are also more difficult to interpret than MLR models [73] and optimisation of network error may get stuck in local minima, rather than reaching the global minima during training [86]. To avoid problems with local minima, networks should be trained multiple times. The increased complexity of ANN models also means that they are susceptible to over-fitting and over-training. Neural networks are trained to minimize prediction error for a training set, but it is prediction of new compounds that is of more interest. Over-training or over-fitting results in an over complicated underlying function, that reduces the error for the training set but leads to poor generalizability for new compounds. The problem of over training can be reduced by early stopping [85], and the use of a selection set, distinct from the training set, to assess generalizability [86]. While training minimizes the true error function, the error for the select set will decrease along with the error for the training set. Over-training is indicated by an increase in the selection set error when the training set error is still decreasing and so can be avoided by selecting networks with balanced errors across the training and selection sets. As the selection set plays a key role in selecting the best network, it is not truly independent from the training process and so an additional independent test set is required [86].

ANNs have shown promise for predicting LC retention times to aid identification of new compounds detected by suspect or non-target screening. Over the past decade, QSRR ANNs have been developed for a number of applications, including: toxicology [88,89], anti-doping [82], wastewater analysis [83,90] and proteomics [91]. As shown in Table 1.7, a range of different methods have been used for descriptor selection including genetic algorithm feature selection, ‘user curation’ where descriptors are selected by the authors and the use of previously published descriptors. The suitability of previously published descriptor sets, has not yet been investigated for energetic materials. Different measurements of accuracy, including root mean square error (RMSE), mean absolute error (MAE) and maximum error (Max E), have been reported, making it difficult to compare accuracy across papers. Additionally, these measures of accuracy give little indication of the confidence associated with a predicted retention time for a new compound. It is also worth noting that all the ANNs described in Table 1.7 were developed on C_{18} stationary phases and so suitability for alternative stationary phases still needs to be investigated.

Table 1.7: Prediction of LC retention time for new compounds using ANNs.

Reference	Application	No. of Analytes	Type of ANN	Column	No. of Descriptors	Descriptor selection	Accuracy
Noorizadeh, 2011 [88]	Toxicology	52	3-layer network	C_{18}	15	Genetic algorithm	RMSE = 0.55
Miller, 2013 [82]	Anti-doping	86	4-layer MLP	C_{18}	18	User curated	Max E < 1 min
Munro, 2015 [83]	Wastewater analysis	166	GRNN	C_{18}	17	User curated	MAE (± SD) = 0.99 ± 1.17 min
Bade, 2015 [90]	Wastewater analysis	550	4-layer MLP	C_{18}	16	Miller et al.	RMSE=1.03 min
Žuvela, 2016 [91]	Proteomics	280	MLP	C_{18}	10	Genetic Algorithm	RMSEP = 9 % of t_R
Mollerup, 2018 [89]	Toxicology	869	4-layer MLP	C_{18}	16	Barron & McEneff	MAE = 0.97 min, 95th percentile = 2.4 min

1.3.2 Evaluating the significance of a match

For forensic analysis, it is essential to evaluate the significance of a match, with or without reference materials, and the probability of misidentification. In 2017 Rochat proposed the use of a confidence scale and identification score for known-unknown compounds in metabolomics using LC-HRMS [92]. The proposed scale included: a general identification category of 1-4, with 1 indicating confirmed and 4 indicating unknown; a chromatography class of A-D, with a matching reference standard required for A, a predicted retention time <10 % for B and <25 % for C and, finally, an identification point level. According to Rochat's scale, 'very strong' identification confidence could be achieved without a reference standard, but for confirmation a reference standard analysed by at least two orthogonal MS analysis, or confirmation by NMR, was required. The scale proposed by Rochat built upon the identification points system set out in the 2002 European Commission Decision [93]. Xu et al. used the same identification points system for their LC-HRMS explosives screening method, with the slight modification of awarding an additional identification point to HRMS ions with a resolution of 60,000-100,000 rather than ~20,000 FWHM [33]. However, Lehotay et al. argued that the identification points system was arbitrary and fundamentally unscientific [94]. Therefore, further evidence is required to support the selection of identification requirements for LC-HRMS analysis. While reference materials will inevitably be required for confirmatory analysis, predicted retention times may be beneficial for excluding false positives with the same elemental composition and prioritising the compounds for which reference materials should be acquired or synthesised.

1.4 Aims and Objectives of the PhD thesis

The overarching aim of this thesis is to investigate the suitability of liquid chromatography coupled to high resolution mass spectrometry for screening and identification of large numbers of known, unknown and suspect energetic materials in forensic casework. As such, the objectives of this thesis are:

A. Develop a liquid chromatography – high resolution mass spectrometry (LC-HRMS) method for broad scope, flexible screening of multiple classes of explosives.
B. Assess qualitative method performance for target compounds, focussing on the requirements for forensic identification of explosives.
C. Assess the ability for the LC-HRMS method to detect new compounds.
D. Use MLR and machine learning techniques for prediction of retention times, to support preliminary identification of new compounds.

1.5 Outline

The main advantages, challenges and proposed solutions to the use of LC-HRMS for screening and identification of energetic materials are summarised in Figure 1.8. LC-HRMS offers the potential to include more explosives in one method, but to do this chromatographic separation and simultaneous ionisation of more explosives with diverse functionalities would be required. Therefore, Chapter 2 will focus on LC-HRMS method development and optimisation for target compounds. HRMS offers greater confidence of identification due to the ability to separate isobaric ions which are indistinguishable by low-resolution MS. However, there is currently a lack of consensus regarding identification criteria for LC-HRMS and limited scientific evidence to demonstrate how low the uncertainty of identification is, in the case of energetic materials. Furthermore, it is important to understand any uncertainties associated with an identification in order to avoid a miscarriage of justice. Therefore, in Chapter 3, the effect of different LC-HRMS identification criteria on selectivity and sensitivity will be investigated and the number of isomers and theoretical elemental compositions indistinguishable by HRMS, from target explosives, will be determined. The third main advantage of LC-HRMS is the ability to perform suspect and biased non-target screening to detect explosives not included in targeted analysis. This presents the challenge of how to optimise and performance test the method for unknown compounds. To address this challenge, the LC-HRMS method was first optimised for a representative set of target

compounds (including nitroaromatics, nitrate esters, nitramines and peroxides), as described in Chapter 2. Method performance was then assessed, as described in Chapter 3, for a larger set of compounds, including compounds for which no optimisation was performed, as would be the case for unknowns. For suspect and non-target compounds there is also the challenge of identification, initially without reference standards. To address this, retention time prediction was investigated in Chapter 4 for the potential to prioritise the purchase/synthesis of reference materials and/or exclude non-explosive compounds. To the authors knowledge, this is the first time the use of LC-HRMS for suspect screening and non-target analysis, in addition to targeted analysis, has been investigated for energetic materials.

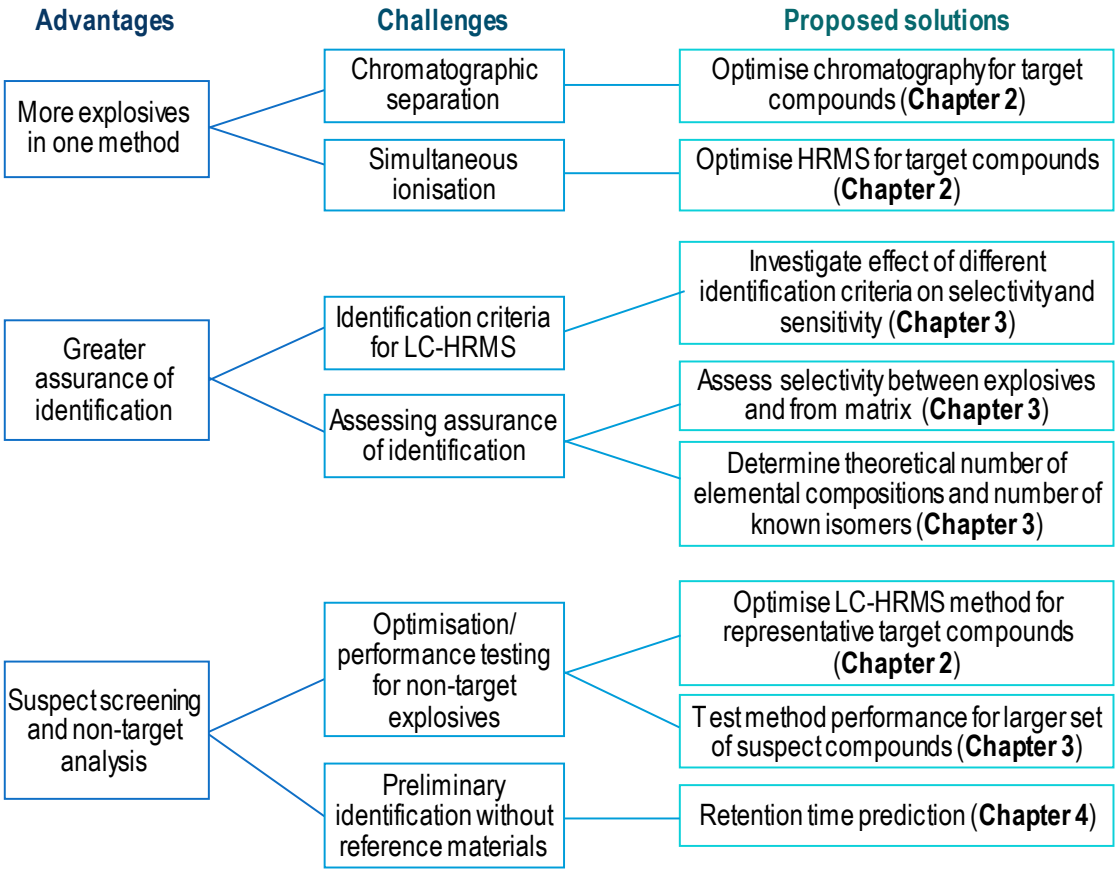


Figure 1.8: Outline of the main advantages, challenges and proposed solutions for the use of LC-HRMS for screening and identification of explosives.

Chapter 2: Multi-residue LC-HRMS method development

2.1 Introduction

High performance liquid chromatography (HPLC) and liquid chromatography coupled to mass spectrometry (LC-MS) has been widely used for screening and confirmatory identification of organic explosives both for environmental monitoring and forensic purposes as reported in Chapter 1 (Table 1.5 and Table 1.6). Compounds of interest vary depending on the application. For example, environmental monitoring focusses on toxic compounds and environmental transformation products, whereas forensic screens are more interested in pre- and post-blast compounds. Many methods have only focused on small numbers of explosives, or even individual explosives [95,96]. This has meant that multiple methods are required to provide comprehensive detection. Methods detecting multiple explosives have often focused on either one class of explosive e.g. nitrate esters [34] or nitroaromatics [42,50,53], or on more environmentally relevant analytes such as the 17 compounds included in the United States Environmental Protection Agency (EPA) methods 8330, 8330 A and 8330 B [57-59,97].

Recently liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) has also been used for forensic screening/identification of explosives [32,33]. The use of LC-HRMS offered the potential to simultaneously screen for a large number of explosives and related compounds that could be confidently identified based on their LC retention time in combination with high resolution, high accuracy *m/z* ratios. With tandem mass spectrometry, the number of explosive transitions that could be measured in the same method without sacrificing the number of data points per chromatographic peak was limited. This was not the case for full-scan high resolution mass spectrometry, as data for all ions is collected in each scan. This means more explosives could be included in one HRMS method. With full-scan HRMS, since data is collected for all ions, suspect or non-target compounds could also be detected through suspect or biased non-target screening of new analytes as and when the need arose. This presents the challenge of how to optimise a method for the detection of non-target analytes that are not selected in advance. To maximise the potential for suspect screening, it is important to optimize a method for target analytes with a range of functionalities to improve the chances of detecting non-target compounds. In Figure 2.1, eighteen initial target analytes of forensic interest are identified including eight nitroaromatics, four nitrate esters, two nitramines, one nitrosamine, two peroxides and tetryl which is both a nitroaromatic and a nitramine. The two main challenges when including

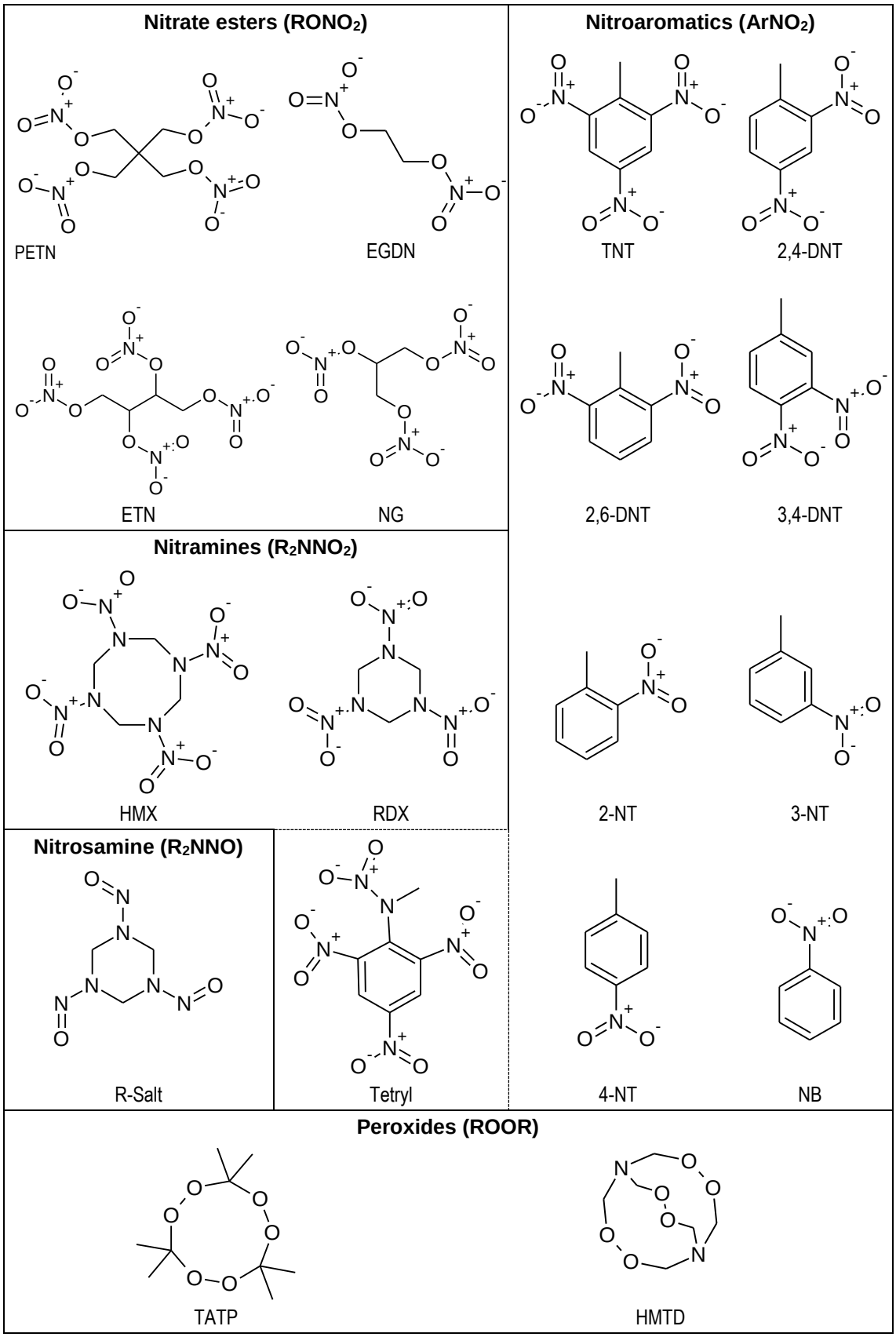


Figure 2.1: Eighteen organic explosives, including sixteen nitro-explosives, grouped based on their chemical structure and identified as initial target compounds for a forensic screen. Note: Tetryl has both nitroaromatic and nitramine functionality

more explosive compounds, with a range of functionalities, in one method are achieving chromatographic separation and sufficient ionization of all compounds.

In order for compounds to be detected by mass spectrometry, first they must be ionised. The two main ionisation sources used for the detection of explosives by LC-MS are electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI), but the ionisation conditions chosen have varied depending upon the explosive analytes included in the method. Since nitrate esters, nitramines and TATP do not readily form molecular ions or (de)protonated molecules, a range of mobile phase additives/buffers have been used to aid adduct formation including ammonium acetate [33,35,39,40,42-45], ammonium chloride [34,35,39], ammonium formate [32,36,37], ammonium nitrate [35] and dichloromethane [38]. Detection of adducts rather than molecular ions can make interpretation more complicated, especially for the identification of non-target compounds where identification of the elemental composition becomes more challenging. Ionisation, and particularly simultaneous ionisation of multiple types of explosives, remains a challenge since different explosives ionise most efficiently under different conditions.

As shown in a 2007 review by Gaurav et al. [56], HPLC has been widely used for explosives. However, baseline separation of the initial target analytes identified in Figure 2.1 is still lacking. Since mass spectrometry can distinguish between ions with different *m/z*, many LC-MS methods have only attempted separation of isomers, which cannot be distinguished by *m/z*. However, for forensic purposes the extra confidence of identification achieved from both chromatographic separation and mass spectra is desirable, particularly when several explosives form common ions. For example, DeTata et al. reported detection of a nitrate ion (*m/z* = 61.9885) for eight different explosive compounds [32].

The main aim of this chapter is to show the development and optimisation of an LC-HRMS method for the screening and identification of multiple classes of organic explosives including the 18 initial target analytes shown in Figure 2.1. Here, a comprehensive chromatographic separation was more thoroughly pursued, in comparison to literature-reported methods through column/stationary phase comparison, gradient and temperature optimisation. A thorough assessment of ionisation parameters, including voltages, temperatures, mobile phase additives and a comparison of ESI and APCI was also completed to optimize ionization for broad screening of organic explosives. Finally, the optimized LC-HRMS method was used for the detection of explosives in latent fingermarks, a novel application for this analytical technique.

2.2 Experimental

2.2.1 Materials and reagents

Unless otherwise stated, all reagents were of analytical or reagent grade and used without further purification. Standard solutions of the 18 initial target explosives: 2,4-dinitrotoluene (2,4-DNT, > 99.9 % purity in methanol), 2,6-dinitrotoluene (2,6-DNT, > 99.9 % purity in methanol), 2-nitrotoluene (2-NT, 99.0 % purity in 50:50 acetonitrile:methanol), 3,4-dinitrotoluene (3,4-DNT, 100 % purity in methanol), 3-nitrotoluene (3-NT, 98.7 % purity in 50:50 acetonitrile:methanol), 4-nitrotoluene (4-NT, 99.2 % purity in 50:50 acetonitrile:methanol), ethylene glycol dinitrate (EGDN, 96.2 % purity in acetonitrile), erythritol tetranitrate (ETN, 99.9 % purity in acetonitrile), hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX, 98.6 % purity in 50:50 acetonitrile:methanol), hexahydro-1,3,5-trinitroso-1,3,5-triazine (R-Salt, 99.8 % purity in acetonitrile), hexamethylene triperoxide diamine (HMTD, 98.4 % purity in acetonitrile), nitrobenzene (NB, 99.8 % purity in methanol), nitroglycerin (NG, 99.4 % purity in purity in 97:3 ethanol:methanol), N-methyl-N,2,4,6-tetranitroaniline (Tetryl, 98.6 % purity in 50:50 acetonitrile:methanol), octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazine (HMX, 98.0 % purity in 50:50 acetonitrile:methanol), pentaerythritol tetranitrate (PETN, 99.3 % purity in methanol), triacetone triperoxide (TATP, 99.9 % in acetonitrile) and trinitrotoluene (TNT, >99.9 % purity in 50:50 acetonitrile:methanol) were purchased from Kinesis Ltd. (St Neots, UK) in either 0.1 or 1 mg mL^{-1} solutions. Also purchased from Kinesis Ltd. were individual standard solutions of diethylene glycol dinitrate (DEGDN, 99.9 % purity in 50:50 acetonitrile:methanol), nitromethane (NM, 100% purity in methanol), trinitrobenzene (TNB, 97.5 % purity in 50:50 acetonitrile:methanol) and picric acid (PA, 99.1% purity in 50:50 acetonitrile:methanol). Solid dimethyl-2,3-dinitrobutane (DMNB, 98 % purity) and 2,3-dinitrotoluene (2,3-DNT, 99.9 % purity) were purchased from Sigma-Aldrich (Gillingham, UK). Individual standard solutions were stored at room temperature, 0-5 °C or -20 °C according to manufacturer's recommendations. Dilutions and mixed standards were prepared using positive displacement pipettes in HPLC grade methanol (Fisher Scientific UK, Loughborough, UK). HPLC grade methanol was also used for mobile phases along with ultrapure water (18.2 MΩ·cm) delivered from a Millipore Milli-Q water ultra-purification system (Millipore, Bedford, MA, USA) and ammonium acetate or ammonium chloride which were purchased from Sigma-Aldrich (Poole, UK). All mobile phases were sonicated for 15 min and filtered before use.

2.2.2 Optimisation of chromatographic separation conditions

An Agilent HP 1100 series chromatograph (Agilent Technologies, Berkshire, UK) with a diode-array detector was initially used for optimisation of chromatographic separation for the 16 nitro-explosives identified as initial target analytes. Column selection experiments were run at 25 °C under isocratic conditions, with an 8 mM ammonium acetate, 60 % methanol, 40 % water mobile phase, at a flow rate of 0.15 mL min^{-1}, unless otherwise stated. Seven columns were investigated: Cogent bidentate C_{18}, Cogent diamond hydride, ACE Excel C_{18} HL, ACE Excel CN-ES, ACE Excel C_{18} Ar, YMC Triart C_{18} and YMC Triart PFP. All columns were obtained from HiChrom ltd. (Berkshire, UK) and further details, including dimensions, are given in Table 2.1.

Table 2.1: Details of columns tested for efficiency and selectivity.

Column	Length (mm)	ID (mm)	Particle size (µm)	Carbon loading (%)	Notes
Cogent bidentate C_{18}	150	2.1	4	16	Two points of attachment to silica
Cogent diamond hydride	150	2.1	4	<2	Silicon hydride surface with very small amount of carbon on surface
ACE Excel C_{18} HL	150	2.1	3	15.5	HL stands for high carbon load
ACE Excel CN-ES	150	2.1	3	12.6	Terminal CN group and 'C18-like' extended spacer (ES)
ACE Excel C_{18} Ar	150	2.1	3	15.5	C18 chain with integral phenyl functionality
YMC Triart C_{18}	150	2	3	20	Highest carbon load
YMC Triart PFP	150	2	3	15	Pentafluorophenyl (PFP)

Gradient separations were later developed on two columns, the YMC Triart PFP and the ACE C_{18} Ar using 10 % CH_3OH, 8 mM ammonium acetate for mobile phase A and 90 % CH_3OH, 8 mM ammonium acetate for mobile phase B. The optimised separation on the YMC Triart PFP column involved a linear gradient of 40-55 % B over 0-50 min and a temperature program starting at 5 °C, increasing to 45 °C at 10 min before decreasing back to 5 °C from 25 min. Whilst the optimised separation on the ACE C_{18} Ar column involved a 40-100 % B linear gradient over 0-30 min at 20 °C.

2.2.3 Optimisation of high resolution mass spectrometry conditions

An Exactive™ mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), with an Ion Max API Source, an Accela UHPLC pump and HTC Pal autosampler (also from Thermo), was used for HRMS analysis. The Ion Max API Source could be used with either an atmospheric pressure chemical ionisation (APCI) or a heated electrospray ionization (HESI-II) probe fitted. A combination of direct infusion, flow injection analysis and LC-MS were used for optimisation of the ion optics and ionisation parameters including ESI spray voltage, HESI heater temperature,

APCI coroner discharge, APCI vaporiser temperature, capillary temperature and the use of mobile phase additives. For experiments using +/- ESI, the optimised conditions were as follows: HESI heater temperature of 300 °C, capillary temperature of 250 °C, sheath gas flow rate of 40, auxiliary gas flow rate of 10, spray voltage of 3 kV, capillary voltage of -25 or +47.5 V, tube lens voltage of -65 or +50 V and a skimmer voltage of -18 or +16 V. For experiments using APCI, the optimised conditions were: vaporizer temperature of 300 °C, capillary temperature of 250 °C, sheath gas flow rate of 50, auxiliary gas flow rate of 5, capillary voltage of -25 or +25 V, tube lens voltage of -55 or +50 V, skimmer voltage of -18 or +18 V and discharge current of 20 or 10 µA. A comparison of APCI and ESI method performance was also performed at 25, 2.5, 0.5 and 0.05 ng on column for the 18 target compounds.

2.2.4 Optimised LC-HRMS conditions

For the optimized method, the APCI conditions listed in Section 2.2.3 were used for ionisation. For chromatographic separation, the ACE C_{18} Ar column (150 × 2.1 mm, 3 µm) with a 1 cm guard column was used. The injection volume was set at 5 µL as an increased injection volume had a detrimental effect on separation. Gradient separation was performed at 0.3 mL min^{-1}, using 10:90 (v/v) methanol:water (A) and 90:10 (v/v) methanol:water (B) with 0.2 mM ammonium chloride in both phases. The optimized gradient was 40-95 % B over 0-15 min, 95-100% B 15-15.5 min, followed by 5 min at 100 % B and a 17.5 min re-equilibration time (half the time required for UV baseline to return to normal at 0.15 mL min^{-1}). The column temperature was maintained at 20 °C using a model 7955 column oven from Jones Chromatography (Hengoed, UK).

Data was acquired using XCalibur software Version 2.2 (Thermo Fisher Scientific), in full scan mode over a mass range of 59.9-625 *m/z*. High resolution (50,000 FWHM) and a mass tolerance of 5 ppm were used in line with previous work. Prior to analysis, the Exactive™ was evaluated and/or calibrated for low mass analysis. For positive mode two additional analytes, N-butylamine (*m/z*=74.09643) and triethylamine (*m/z*=102.12773), were added to the standard positive ion calibration solution which contained caffeine, MRFA (met-arg-phe-ala) and Ultramark 1621 (a mixture of fluorinated phosphazines). A custom calibration solution containing sodium taurocholic acid (*m/z*=514.28441), sodium dodecyl sulphate (*m/z*=265.14791), phthalic acid (*m/z*=165.01933), benzoic acid (*m/z*=121.02950), glycolic acid (*m/z*=75.00877) and acetic acid (*m/z*=59.01385), all purchased from Sigma-Aldrich, was used for negative mode mass calibration.

2.2.5 Fingermark deposition and extraction procedures

As recommended by Sears *et al*. [98], natural fingermarks, collected from donors during their normal daily routine, were used for all fingermark experiments. In order to produce natural fingermarks, for at least 30 minutes prior to sample collection, hands were not washed, not wiped deliberately across the face nor used to apply cosmetics. Fingermarks were collected and extracted based on a method used by Menzies for LC-MS analysis of drugs in fingermarks (unpublished work). Natural fingermarks were deposited onto 15 mm diameter glass coverslips (VWR International, Lutterworth, UK) after rubbing fingertips together to evenly distribute any residues. For recovery experiments, 6 donors (3 male, 3 female) were used and 0.5 µg of each analyte (100 µL of a 5 µg mL^{-1} mixed standard solution) was added to each fingermark and allowed to dry for 10 min before extraction. Fingermarks were then extracted in 15 mL scintillation vials with 0.5 mL of a 50:50 methanol:water solution using a KS 260 basic shaker (Staufen, Germany) at 300 rpm for 10 min.

C4, a commercial explosive containing RDX, was acquired, stored and handled by licenced technicians from Precision Energetics (Somerset, UK) and EPC UK Ltd. (Essex, UK) in compliance with the Explosives Act 1875 [8] and the Explosives Regulations 2014 [14]. As with blank and spiked fingermarks, hands were not washed, wiped deliberately across the face or used to apply cosmetics for 30 minutes prior to sample collection. After handling, a fingermark depletion series (n=60) was produced by a single donor and a subset of marks were extracted and analysed as above.

Ethical approval for the chemical analysis of energetic materials in fingermarks was granted by the King's College London Biomedical Sciences, Medicine, Dentistry and Natural & Mathematical Sciences Research Ethics Sub-Committee (BDM RESC, Study reference: HR-15/16-1962).

2.2.6 Data analysis

HPLC-UV data were acquired using Agilent ChemStation B.02.01 software. ChemStation was also used to determine retention times (t_R), peak width at half height ($W_{0.5}$) and peak symmetry (A_s), which was defined as simply peak area before apex, divided by peak area after apex, unless inflection points were found [99]. Retention factor (k), selectivity factor (α) and resolution (R_s) were then calculated using Equation 2.1, Equation 2.2 and Equation 2.3. Chromatograms were plotted in Excel after extracting the raw data from ChemStation.

Equation 2.1: Retention factor

$$k = \frac{t_R - t_0}{t_R}$$

Equation 2.2: Selectivity factor

$$\alpha = \frac{k_2}{k_1}$$

Equation 2.3: Resolution using peak width at half height

$$R_s = \frac{2(t_2 - t_1)}{1.7(W_{0.5,1} + W_{0.5,2})}$$

LC-HRMS data were acquired using Thermo Xcalibur version 2.2. Xcalibur Qual Browser, with a 5 ppm mass accuracy threshold from the exact *m/z*, was used to produce extracted ion chromatograms (EICs). Seven-point gaussian smoothing was used prior to integration for determination of peak height and/or area. The raw data for the EICs were normalized in excel to the largest and smallest intensity across the full mass chromatogram. Excel was used to plot all graphs. However, Minitab 18 was used to determine the line of best fit for the scatter plot in Figure 2.10 to allow consideration of variance in both x and y values.

2.3 Results and discussion

2.3.1 Optimisation of chromatography conditions

Initial chromatographic optimisation was performed using HPLC-UV rather than HPLC-HRMS. In addition to cost savings, using a UV detector allowed optimisation of chromatography without having to account for ionization challenges. The intensity of detector response for different compounds was also more similar (for the nitro-explosives that could be detected by UV) than when an HRMS detector was used. This allowed the resolution between two similar sized peaks to be calculated more easily. That said, since the peroxides (e.g. TATP and HMTD) do not contain a chromophore, they do not absorb UV light and so could not be detected by HPLC-UV. This meant that initial chromatographic optimization was only performed for the 16 nitro-explosives identified as initial target compounds and separation of the two peroxides was only assessed after optimisation of HRMS conditions.

2.3.1.1 Separation on YMC Triart PFP column

Initial work focussed on improving an existing chromatographic separation developed by L. Salvia using a pentafluorophenyl (PFP) column during previous work [100]. When comparing seven columns (six with a C_{18} stationary phase and one with a PFP phase), as part of an MSc research project, Salvia found the YMC Triart PFP column to give the best combined selectivity and efficiency for organic explosives and a gradient separation of 45-55 % B over 50 min, at 20 °C, was developed on this column. However, the desired baseline resolution was not achieved for all compounds and three pairs, HMX/R-Salt, NG/NB and 2,6-/2,4-DNT were unresolved. As baseline resolution was desirable for identification, here attempts were made to improve separation on the YMC Triart PFP column through the use of alternative gradients and temperature.

2.3.1.1.1 Gradient separation

A selection of the gradients investigated are shown in Figure 2.2, along with the gradient previously developed (45-55 % B over 0-50 min). Separation was improved by using a 40-55 % B gradient over 50 min.

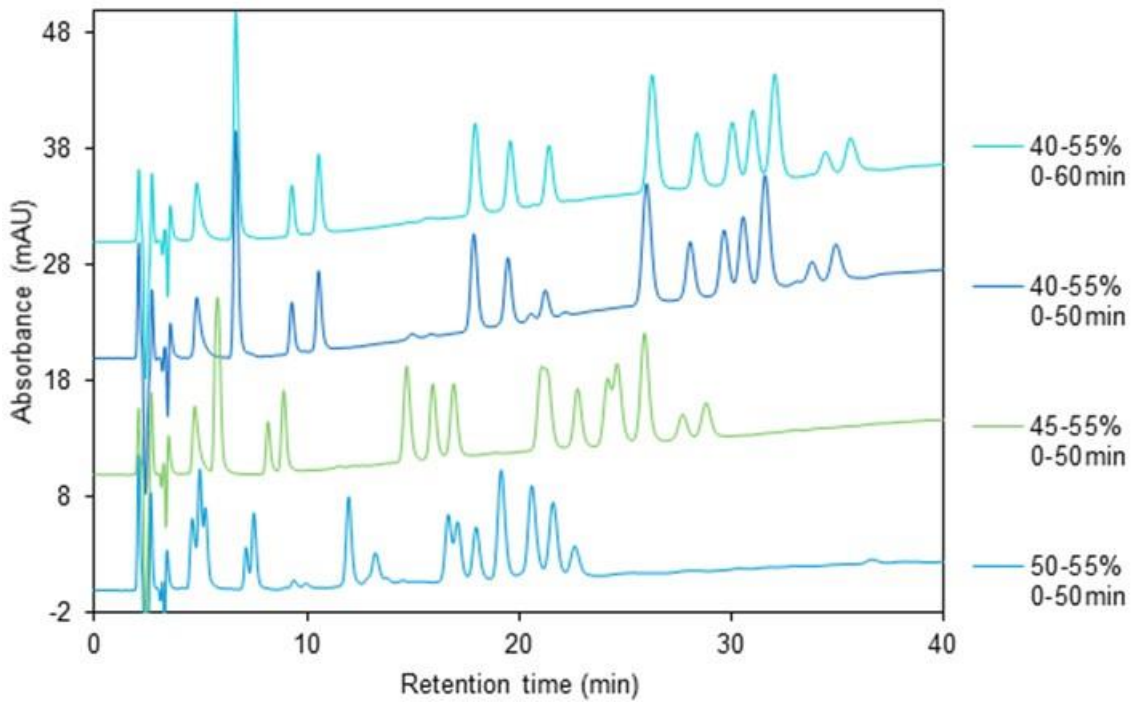


Figure 2.2: Gradient separations of 16 target analytes on a YMC Triart PFP column using: 50-55% B over 0-50 min; 45-55% B over 0-50 min; 40-55% B over 0-50 min and 40-55% B over 0-60 min. All were at 20 °C.

2.3.1.1.2 Effect of temperature on gradient separation

As shown by the van't Hoff equation (Equation 2.4), which is defined under isocratic conditions, retention factor (*k*) is temperature dependent.

Equation 2.4: Van't Hoff equation

$$lnk = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$

To investigate the effect of temperature on this gradient separation, the optimised mobile phase gradient (40-55% B over 0-50 min) was run over a range of temperatures from 5 – 45 °C (Figure 2.3). Separation of the first two peaks, R-Salt and HMX, was best at 5 °C. However, at 5 °C the efficiency and separation of the later eluting analytes, which was best at 20 °C, was lost. In contrast, separation of the isomers 2,4- and 2,6-DNT was best at 45 °C and the group of 4 analytes in the middle of the run (NB, NG, TNT and Tetryl) were separated best at either 5 or 35 °C. So in conclusion, different temperatures resulted in the optimum separation for different regions of the chromatogram and no one temperature achieved separation of all 16 initial target analytes. As a result of this, thermal programs were investigated.

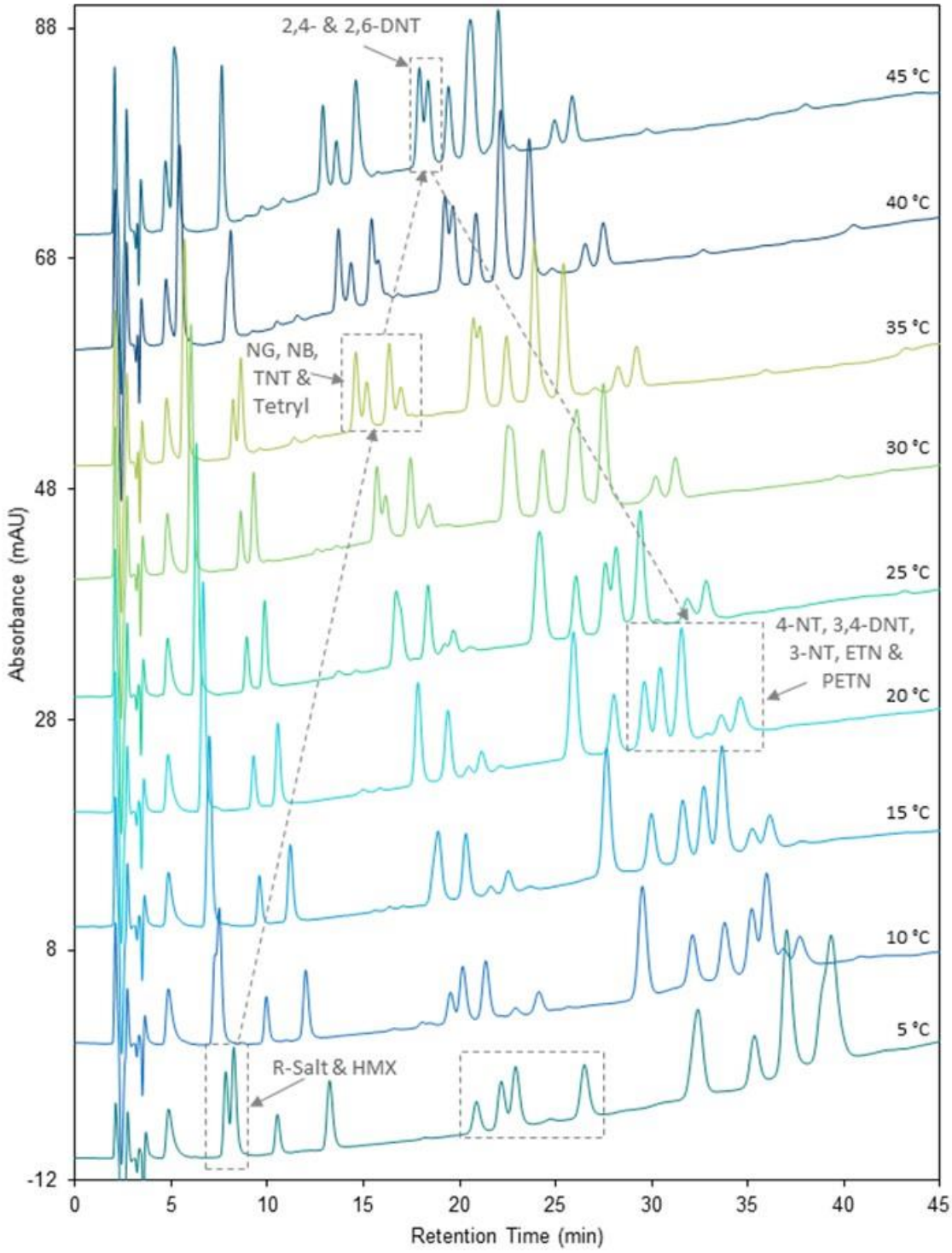


Figure 2.3: The effect of temperature on separation of 16 target analytes using the YMC Triart PFP column and a gradient of 42-54 % CH_3OH over 0-50 min. Dashed boxes indicate areas of best separation.

2.3.1.1.3 Optimised gradient with temperature program

A thermal program, starting at 5 °C before increasing to 45 °C from 10 min and then decreasing back to 5 °C from 25 min, gave the best separation. As shown by the dashed line in Figure 2.4, once heating began at 10 min, it took 7 min to reach 45 °C and cooling took 20 min. Separation

of PA, 2,3-DNT and all initial target analytes except for the isomers 2,4- and 2,6-DNT, was achieved (Figure 2.4). In order to further evaluate the separation, retention times and their relative standard deviation (RSD) were measured over 10 repeat injections. These are given in Table 2.2, along with calculated retention factors (k), peak symmetry (A_s), selectivity factor (α) and resolution (R_s). Unlike retention time (t_R), the retention factor takes into consideration the void time (t_0) and so allows comparison across columns with different dimensions. Peak symmetry refers to the peak shape and for an ideal Gaussian peak $A_s = 1$; values of $A_s < 1$ indicate peak tailing and $A_s > 1$ indicate peak fronting. Selectivity factor refers to the difference in retention factor of two adjacent peaks, with values of 1.00 indicating identical retention factors. Resolution is perhaps the best measure of separation as it takes into account peak width, or in this case peak width at half height ($W_{0.5}$), as well as the two retention times (t_1 and t_2) and for Gaussian peaks $R_s > 1.5$ indicates baseline resolution. In addition to 2,4- and 2,6-DNT, R-Salt/HMX, NG/NB, 4-NT/3,4-DNT, 3,4-DNT/3-NT and ETN/PETN were also not fully baseline resolved with a resolution (R_s) less than 1.5 as shown in Table 2.2. As a result of this, alternative columns were investigated.

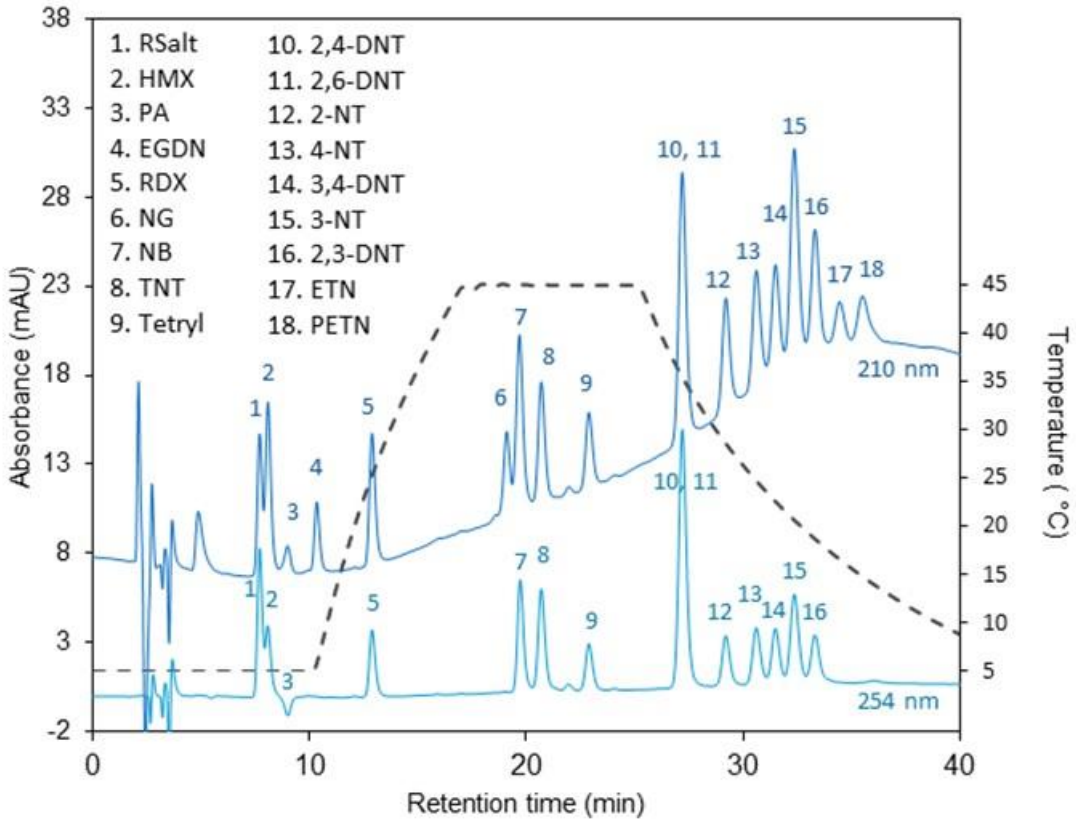


Figure 2.4: Optimised separation including two additional target analytes using the YMC Triart PFP column, a gradient of 42-54 % CH_3OH over 0-50 min and a temperature program of 5-45 °C mid-run. The measured oven temperature throughout the run is indicated by the dashed line.

Table 2.2: Retention, symmetry, selectivity and resolution of optimised separation on the YMC Triart PFP column.

Analyte	Retention time (min)	RSD (%)	Retention factor (k)	Peak symmetry (A_s)	Selectivity factor (α)	Resolution (R_s)
NM	3.80	0.04	0.82	0.81	1.82	5.54
Sys peak†	4.83	0.04	1.32	0.50	1.27	2.14
R-Salt*	7.94	0.13	2.81	0.79	1.64	5.70
HMX*	8.49	0.16	3.08	0.77	1.07	1.07
PA	8.94	0.18	3.30	0.70	1.05	0.82
EGDN*	10.53	0.09	4.06	0.74	1.18	3.10
RDX*	13.42	0.16	5.45	0.76	1.28	5.55
DEGDN	14.20	0.10	5.82	0.82	1.06	1.44
NG*	19.85	0.09	8.54	0.95	1.40	10.40
NB*	20.47	0.13	8.84	0.85	1.03	1.13
TNT*	21.41	0.09	9.29	0.82	1.05	1.63
Tetryl*	23.82	0.12	10.44	0.85	1.11	4.09
2,4 DNT*	27.45	0.01	12.19	0.82	1.15	6.00
2,6 DNT*	27.47	0.19	12.20	0.81	1.00	0.03
2-NT*	29.07	0.13	12.97	0.80	1.06	2.55
4-NT*	30.26	0.12	13.54	0.82	1.04	1.79
3,4-DNT*	31.23	0.12	14.00	0.82	1.03	1.41
3-NT*	31.86	0.11	14.31	0.85	1.02	0.88
2,3-DNT	32.97	0.11	14.84	0.83	1.04	1.50
ETN*	34.51	0.06	15.58	0.75	1.05	2.12
PETN*	35.50	0.06	16.05	0.79	1.03	1.30

*Initial target analytes
†Unidentified system peak present in all injections at 210 nm.

2.3.1.2 Column Selection

Key details of the six columns investigated alongside the YMC Triart PFP column, in order to compare efficiency and selectivity, are given in Table 2.1. A range of different phases were investigated including C_{18}, bidentate C_{18}, a cyano phase with a C_{18}-like extended spacer (CN-ES), a C_{18} phase with phenyl functionality (C_{18} Ar) and a diamond hydride phase. Column dimensions and particle sizes were selected that were as similar as possible to the PFP column, to aid comparison of the different phases. However, whilst all columns were available with 150 mm length, for some the closest internal diameter (ID) was 2.1 mm (rather than 2 mm) and the smallest particle size available for the two cogent columns was 4 µm (rather than 3 µm). Using a larger particle size results in more band broadening and a less efficient separation. This is due to larger multiple path (*A*) and mass transfer (*C*) terms resulting in a larger plate height (H) as described by the van Deemter equation (Equation 1.3), where u_x is the linear velocity and B the longitudinal diffusion constant. Using a smaller particle size leads to increased backpressure and for sub 2 µm particles, sharper peaks may prevent acquisition of enough MS scans to fully define

a chromatographic peak.

2.3.1.2.1 Efficiency/van Deemter plots

As efficiency affects peak width and hence separation, the efficiency of the columns was compared by constructing van Deemter plots with RDX as a probe (Figure 2.5). The most efficient columns have the smallest plate heights and so are at the bottom of the plot. Perhaps unsurprisingly, the two least efficient columns were the two with the largest particle sizes, the cogent diamond hydride and bidentate C_{18} columns. The diamond hydride column had particularly poor efficiency which may be due to the mobile phase system used for comparison purposes, as this column has a very different phase to the other columns tested. Alternatively, this phase may just be better suited to separation of more polar analytes. At the other end of the spectrum, the CN-ES column was the most efficient followed by the PFP and then the C_{18} Ar. The CN-ES column also had a very flat C term. With good efficiency maintained at higher flow rates, faster analysis may be possible.

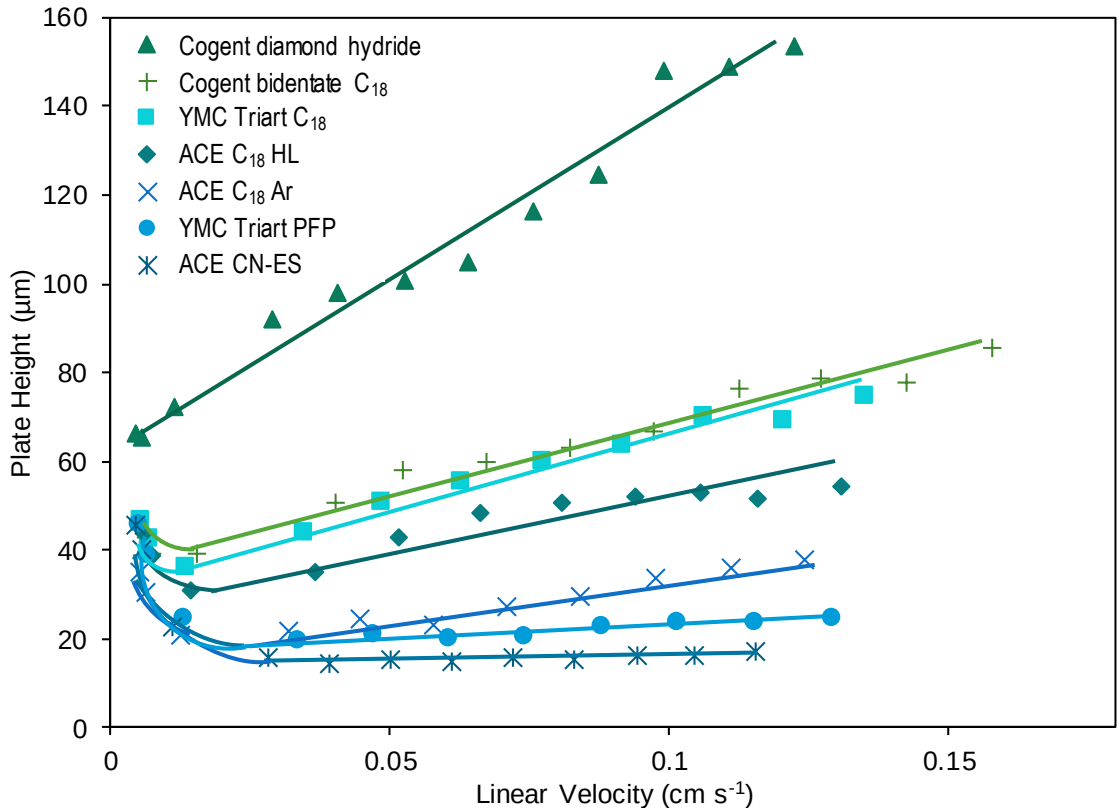


Figure 2.5: Overlaid van Deemter plots, with RDX as a probe, for the comparison of column efficiencies.

2.3.1.2.2 Separation potential

Along with good efficiency it is important to have good selectivity in order to achieve separation. The separation potential of the columns was investigated through comparison of an isocratic separation with 60 % methanol and 8 mM ammonium acetate (Figure 2.6). At this stage, the separation potential (i.e. how many distinct peaks could be detected for the mix of 16 target explosives) was of more interest than the retention order. Very poor retention was seen on the diamond hydride column, again this may be because the column is unsuitable for the separation of nitro-explosives or that an alternative mobile phase system would be required.

The increased retention of RDX on the CN-ES column may explain why this column had the best efficiency according to the van Deemter plot in Figure 2.5, since efficiency (N) is directly proportional to retention time, as well as inversely proportional to peak width. The best selectivity was seen with the ACE C_{18} Ar column which also had good efficiency and so was chosen to take forward for development of a separation.

2.3.1.3 Separation on ACE C_{18} Ar column

2.3.1.3.1 Isocratic separation and the effect of temperature

To investigate the effect of temperature on the ACE C_{18} Ar column, the isocratic separation was run over a range of temperatures from 5 – 45 °C and a van't Hoff plot showing the relationship between *1/T* and *lnk* was produced (Figure 2.7). Using the van't Hoff equation (Equation 2.4), the enthalpy change, ΔH, was then calculated from the slope of the van't Hoff plot which is $-\Delta H/R$, where R is the gas constant. For all analytes, ΔH values were negative, indicating exothermic interactions with the stationary phase. Similar ΔH values, as seen for NG and NB, indicated similar temperature selectivity. This was not the case for all explosives, with the largest difference in ΔH values, and hence temperature selectivity, seen for EGDN (ΔH = -7.62 kJ mol^{-1}) and Tetryl (ΔH = -16.1 kJ mol^{-1}).

Retention times for all the initial target analytes decreased with increased temperature. The last peak (PETN) reduced from 28.2 min at 5 °C to 15.3 min at 45 °C, as expected for exothermic interactions. Efficiency (N) also decreased with increased temperature (for PETN N =16200 at 5 °C and N =11500 at 45 °C), which is perhaps less expected as increased temperature typically reduces the effect of mass transfer and hence band broadening. Selectivity changes were observed for EGDN/RDX, 3-NT/2,6-DNT and Tetryl/ETN, with the pairs coeluting before switching

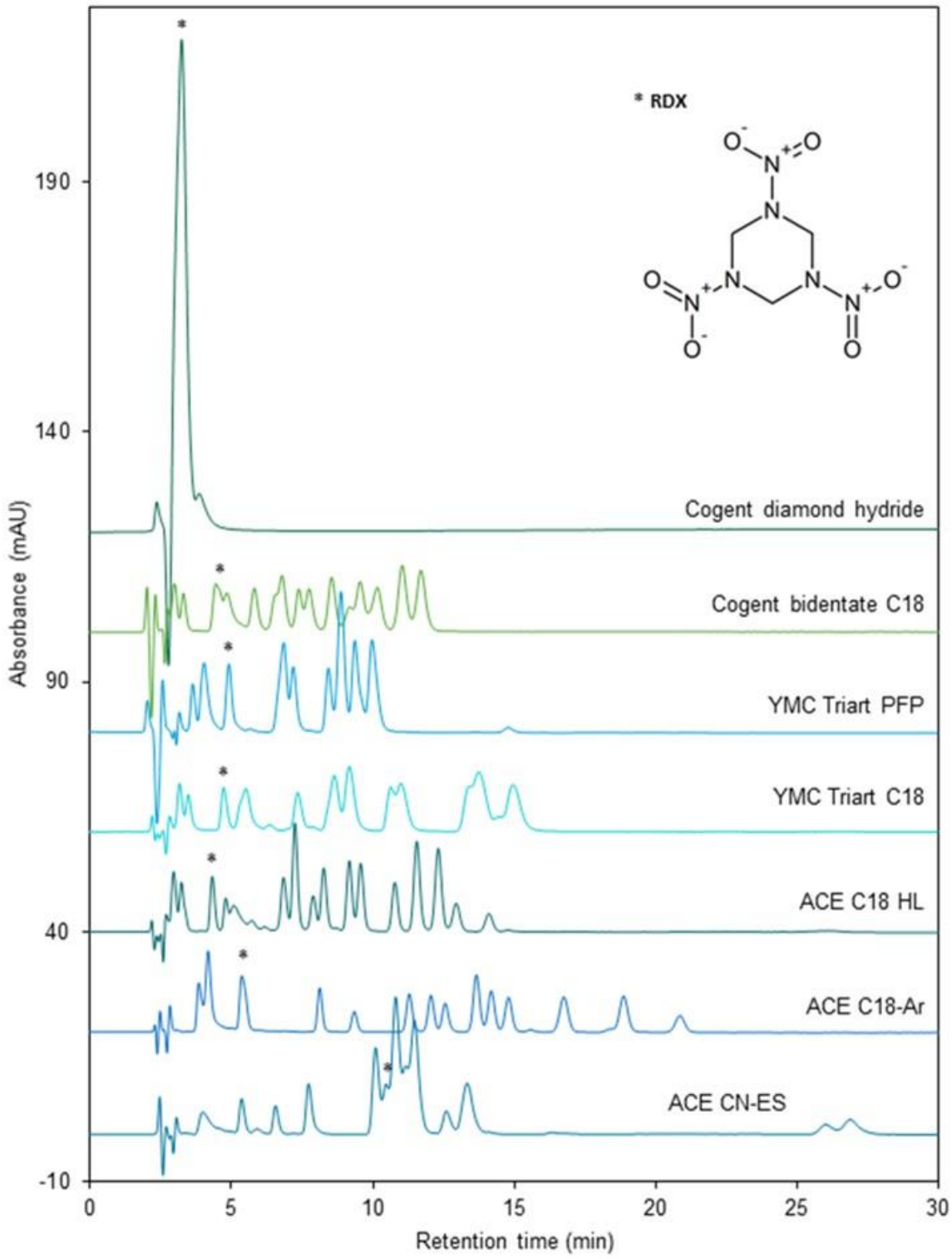
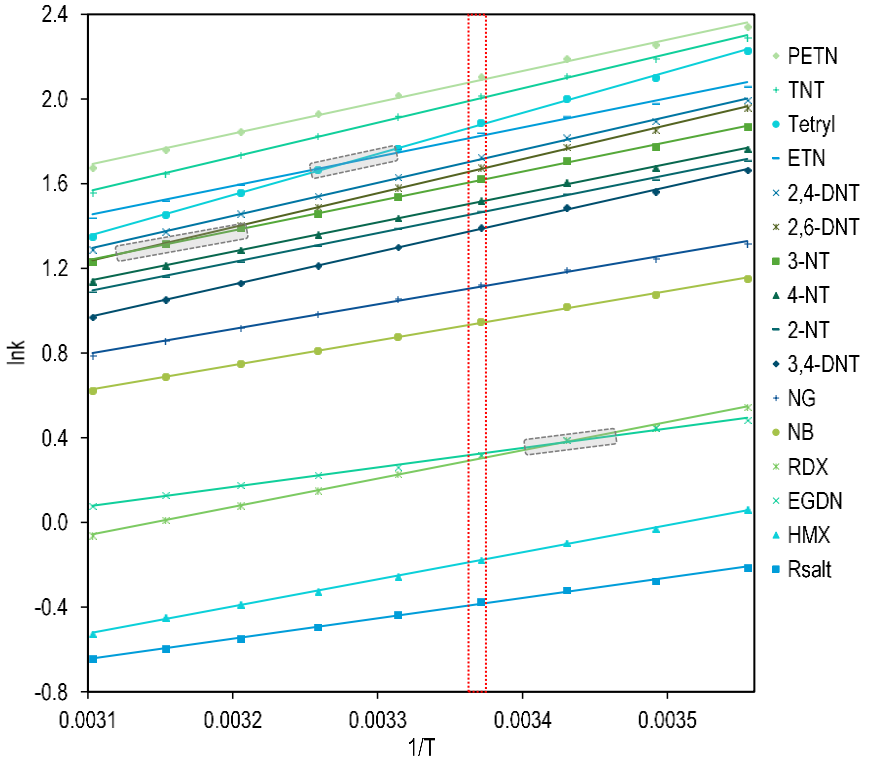


Figure 2.6: Isocratic separations of 16 target analytes at 210 nm. The position of RDX in each chromatogram is indicated by an asterisk.

elution order as temperature is altered due to differences in *ΔH* values. The best separation was achieved at 20 °C with all analytes baseline resolved apart from RDX and EGDN which coeluted and R-Salt which was not base-line resolved from the unidentified system peak present in all chromatograms at 210 nm.



	R^2	*m*	*ΔH (kJ mol⁻¹)*
PETN	0.996	1480	-12.3
TNT	0.999	1620	-13.5
Tetryl	0.999	1940	-16.1
ETN	0.996	1380	-11.5
2,4-DNT	0.999	1560	-13.0
2,6-DNT	0.999	1610	-13.4
3-NT	0.999	1390	-11.6
4-NT	0.999	1380	-11.5
2-NT	0.999	1370	-11.4
3,4-DNT	0.999	1540	-12.8
NG	0.997	1170	-9.71
NB	0.999	1170	-9.71
RDX	0.998	1330	-11.1
EGDN	0.996	917	-7.62
HMX	0.999	1280	-10.6
Rsalt	0.998	962	-8.00

Figure 2.7: Van't Hoff plot for 16 initial target analytes on a C_{18}Ar column over a temperature range of 5-45 °C. Areas of selectivity change are highlighted in grey, the best separation is indicated by a dashed red rectangle and R^2 values, slope (m) and ΔH values are shown on the right to 3 significant figures.

2.3.1.3.2 Gradient optimisation

Gradient elution was investigated in order to improve separation (Figure 2.8). Of the 20 min gradients ending with 70 % B (Figure 2.8 A), starting at 40 % B gave the best separation in the first half of the chromatogram (between R-Salt and the system peak and EGDN/RDX). However, separation of the second half of the chromatogram was best when starting at 60 % B. Therefore, both steeper gradients and multi-step gradients were attempted (Figure 2.8 B & C). The baseline for 210 nm was adversely affected by the use of a multi-step gradient, so single-step gradients were favoured for UV analysis. The final optimised gradient was 40-100 % B over 0-30 min.

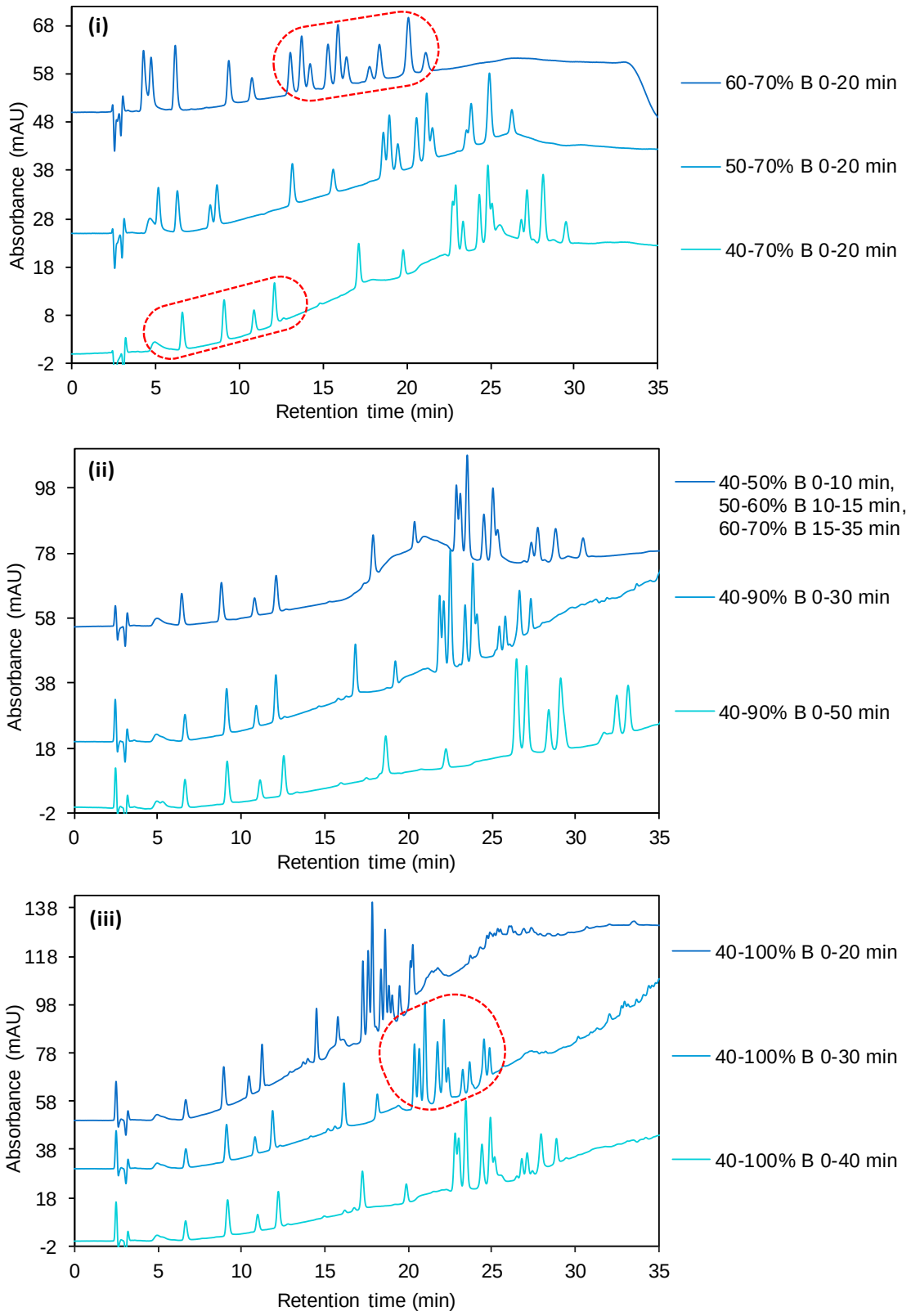


Figure 2.8: Gradient method development on ACE C_{18} Ar column (λ = 210 nm). (i) Shows the effect of different starting mobile phase composition. (ii) Shows multi and singe step gradients starting with 40 % B. (iii) shows the effect of gradient slope with 40-100 % B.

2.3.1.3.3 Optimised gradient

The optimised separation of the 16 initial target analytes, along with NM, DEGDN, PA and DMNB is shown in Figure 2.9 at both 210 and 254 nm. As in section 2.3.1.1.3, average retention time, relative standard deviation (RSD), retention factor (k), peak symmetry (A_s), selectivity factor (α) and resolution (R_s) were also determined for the optimised separation on the ACE C_{18} Ar column (Table 2.3). All analytes were separated, although 2,4- and 2,6-DNT were still not baseline resolved.

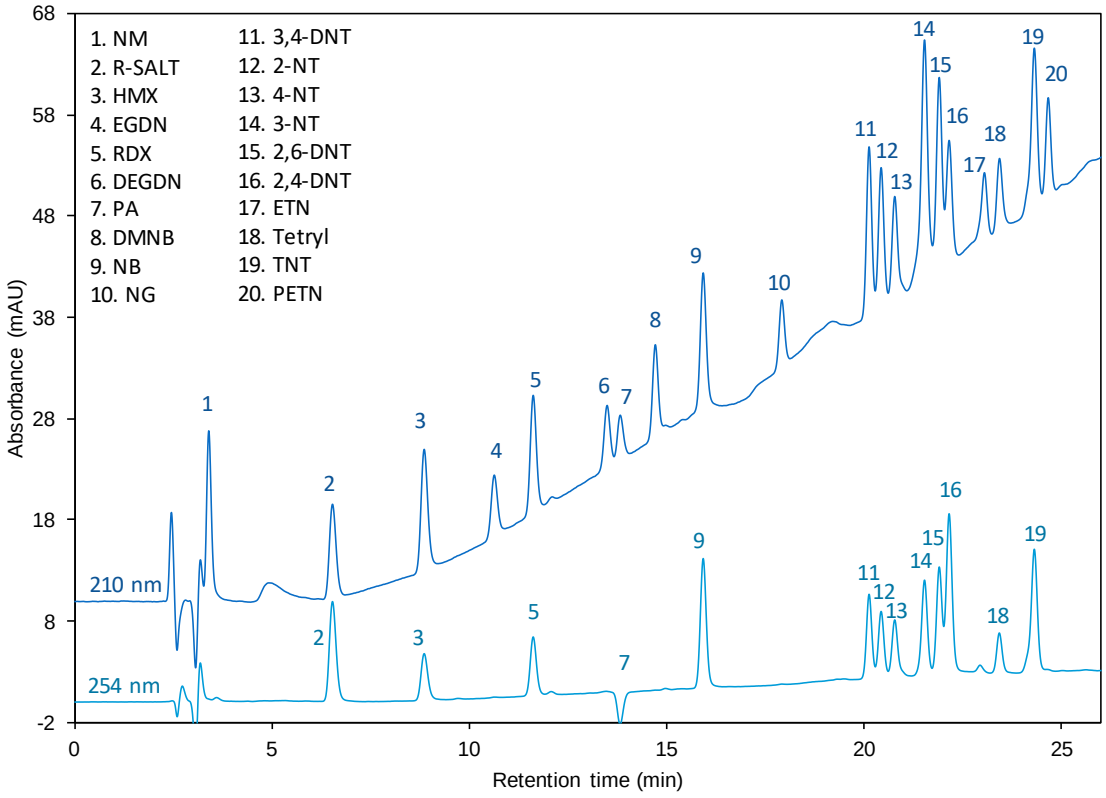


Figure 2.9: Optimised separation including some additional target analytes on the ACE C_{18} Ar column using a linear gradient, 40-100 % B over 0-30 min, at 20 °C.

Table 2.3: Retention, symmetry, selectivity and resolution of optimised separation on the ACE C_{18} Ar column.

Analyte	Retention time (min)	RSD (%)	Retention factor (k)	Peak symmetry (A_s)	Selectivity factor (α)	Resolution (R_s)
NM	3.41	0.11	0.31	0.76	1.06	0.98
sys peak†	4.93	0.15	0.89	0.43	1.45	2.19
R-Salt*	6.57	0.20	1.51	0.90	1.33	2.18
HMX*	8.96	0.42	2.40	0.87	1.36	6.86
EGDN*	10.69	0.19	3.09	0.98	1.20	5.27
RDX*	11.71	0.28	3.46	0.91	1.09	3.06
DEGDN	13.57	0.24	4.18	0.97	1.16	6.06
PA	13.96	0.36	4.31	0.89	1.02	1.08

Table 2.3 (Continued): Retention, symmetry, selectivity and resolution of optimised separation on the ACE C_{18} Ar column.

Analyte	Retention time (min)	RSD (%)	Retention factor (k)	Peak symmetry (A_s)	Selectivity factor (α)	Resolution (R_s)
DMNB	14.80	0.25	4.65	0.94	1.06	2.91
NB*	16.00	0.21	5.12	0.86	1.08	4.01
NG*	18.02	0.25	5.88	0.98	1.13	6.91
3,4-DNT*	20.24	0.25	6.73	0.92	1.12	8.26
2-NT*	20.53	0.21	6.85	0.96	1.02	1.12
4-NT*	20.88	0.21	6.98	0.90	1.02	1.20
3-NT*	21.63	0.21	7.27	1.01	1.04	2.38
2,6-DNT*	22.01	0.19	7.41	1.01	1.02	1.18
2,4-DNT*	22.26	0.21	7.51	0.95	1.01	0.87
ETN*	23.17	0.24	7.85	1.13	1.04	3.17
Tetryl*	23.56	0.22	8.00	0.88	1.02	1.37
TNT*	24.43	0.20	8.34	1.13	1.04	2.97
PETN*	24.80	0.23	8.48	1.06	1.01	1.24

*Initial target analytes
†Unidentified system peak present in all injections at 210 nm.

2.3.1.4 Comparison of two optimised LC methods

The main advantage to the optimised separation on the ACE C_{18} Ar column over the YMC Triart PFP column was separation of the 2,4- and 2,6-DNT isomers as the resolution achieved using the ACE C_{18} Ar column (R_s = 0.87, Table 2.3) was much improved compared to the YMC Triart PFP separation (R_s = 0.03, Table 2.2). Resolution between R-Salt and HMX was also improved (R_s = 6.86 with C_{18} Ar compared to R_s =1.07 with PFP), as was resolution between 3,4-DNT and the adjacent NT peak (R_s = 1.12 between 3,4-DNT and 2-NT with C_{18} Ar compared to R_s = 0.88 between 3,4-DNT and 3-NT with PFP. One pair of peaks where resolution was slightly better with the PFP method was 2- and 4-NT (R_s = 1.20 with C_{18} Ar compared to R_s =1.79 with PFP). Another advantage of the ACE C_{18} Ar separation was that it was simpler, as unlike with the PFP separation thermal programming was not required.

Several HPLC methods for the identification of explosives use multiple columns with orthogonal selectivity in order to achieve confirmation [58,59]. Given that there was a change in retention order between the two optimised methods, their orthogonality was investigated by plotting the retention factor on the PFP column (k_{PFP}) against the retention factor on the C_{18} Ar column (k_{C18Ar}), Figure 2.10. Positive correlation was seen, which was unsurprising given that both phases had aromatic functionality. However, clearly the data did not fit to the line of y=x, indicating that the two methods differed. For all analytes, except picric acid (PA), retention factors were larger using

the PFP method. This may have been due to the more polar mobile phase (lower percentage of methanol) used in the PFP method which would result in increased retention of less-polar compounds. Interestingly, the three analytes that were relatively more retained on the C_{18} Ar column: PA (2,4,6-trinitrophenol), TNT (2,4,6-trinitrotoluene) and Tetryl (N-methyl-N,2,4,6-trinitroaniline), all contain a 2,4,6-trinitro substituted benzene ring. This may be explained by the three electron-withdrawing nitro groups of the 2,4,6-trinitro substituted benzene ring resulting in the formation of stronger intermolecular π-π interactions with phenyl groups than pentafluorophenyl (PFP) groups which have five electronegative fluorine substituents, which are also electron-withdrawing leading to increased repulsion when π-π stacking two electron deficient rings.

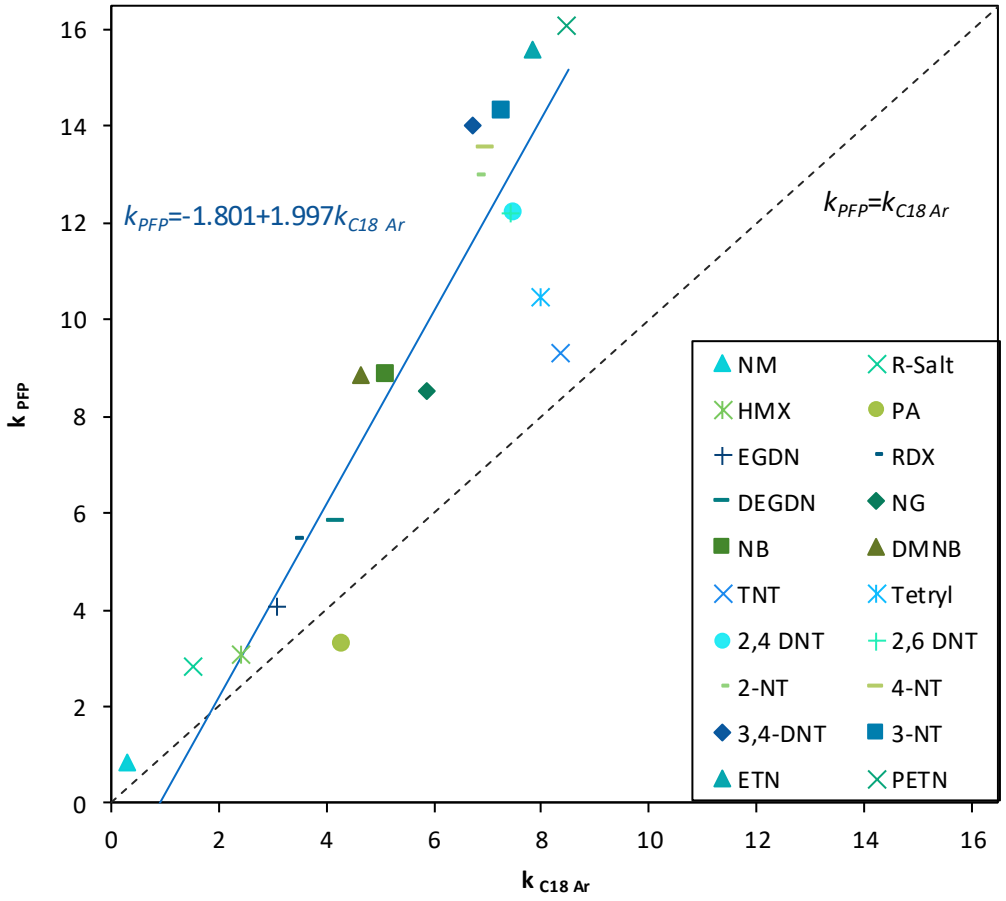


Figure 2.10: Scatter plot showing positive correlation between retention factor on the PFP column (k_{PFP}) and retention factor on the C_{18}Ar column (k_{C18AR}) for target analytes, with the orthogonal regression line and equation shown in blue.

2.3.1.5 Increasing flow rate for LC-HRMS

Including a wash step and re-equilibration, the total runtime for the separation at 0.15 mL min^{-1} was 75 min. While relatively low throughput analysis means runtime is less important than confidence of identification for the forensic detection of explosives, a faster runtime would still be desirable. In order to shorten the total runtime, the effect of increased flow rate was investigated (Figure 2.11). This had the advantage of also shortening the wash and re-equilibration steps, doubling the flow rate to 0.3 mL min^{-1} also halved the total runtime to 37.5 min. However, due to an increased back pressure of 450 bar, it was only possible to use higher flow rates with the ultra-high performance liquid chromatograph (UHPLC) attached to the HRMS and not for the HPLC-UV system used. Therefore, using the instruments available, the effect of increased flow rate on separation could only be assessed for those analytes which were detected by HRMS. Baseline resolution between TNT and PETN was lost when the method was transferred from the HPLC-UV system to the LC-HRMS system and a guard column was added. For all detected analytes except TNT and PETN, separation was maintained with increased flow rate. The considerably faster analysis time justified the slight loss in separation at 0.3 mL min^{-1}, especially whilst optimising HRMS conditions.

Even with HRMS, some of the overlaid extracted ion chromatograms (EICs) used in Figure 2.11 had multiple peaks of different intensities, indicating common ions. For example, nitrate esters NG, ETN and PETN all produced peaks for the nitrate ion extracted mass (61.98837 ± 5 ppm), with the largest being due to PETN. Perhaps more surprisingly, the largest peak seen for the DNT $[M-H]^-$ ion extracted mass (181.02548 ± 5 ppm) was actually due to a fragment of Tetryl rather than any of the DNT isomers, demonstrating the importance of LC separation in the detection of organic explosives.

2.3.2 Optimisation of high resolution mass spectrometry

Both electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI) conditions were optimised before the performance of the two ionisation techniques were compared for the 18 initial target compounds.

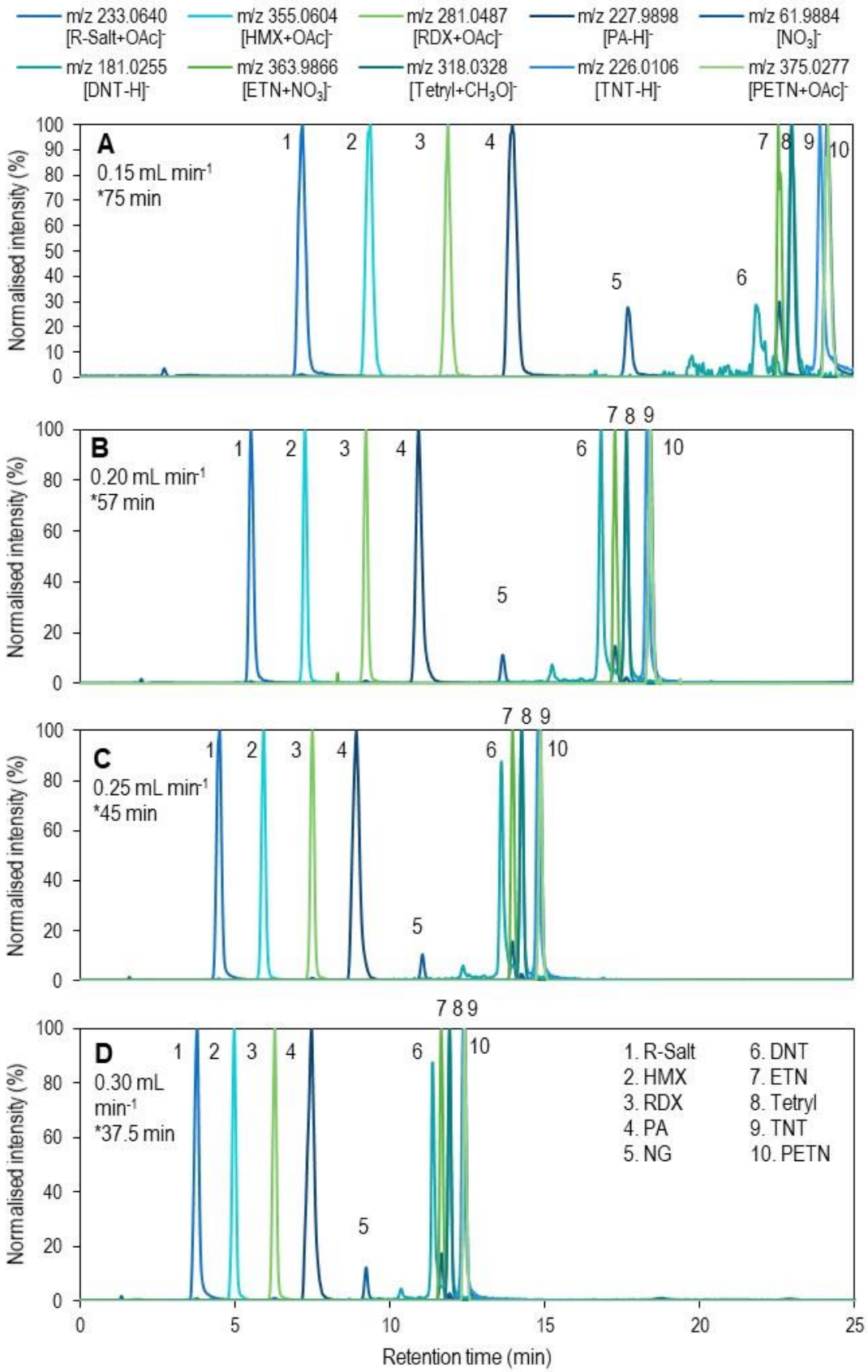


Figure 2.11: Overlaid LC-(ESI)-HRMS extracted ion chromatograms for target analytes at 0.15, 0.2, 0.25 and 0.3 mL min^{-1}. The total runtime including re-equilibration is also given and indicated by an asterisk.

2.3.2.1 Optimisation of electrospray ionisation – high resolution mass spectrometry conditions

2.3.2.1.1 Optimisation of ESI-HRMS parameters using ammonium acetate mobile phase

Heated electrospray ionisation mass spectrometry (HESI-MS) parameters were initially optimised with the 8 mM ammonium acetate, methanol:water mobile phase used for optimisation of chromatography in Section 2.3.1 and in existing LC-MS/MS methods at the Forensic Explosives Laboratory (unpublished work) where the ammonium acetate was used to form acetate adducts of NG, RDX, HMX and PETN and ammonium adducts of TATP. As ammonium acetate was also used for other research projects regularly using the same LC-HRMS instrument for the detection of pharmaceuticals [83], another reason for starting with this mobile phase additive was that it minimised the time required to equilibrate the system between users.

Ion optic parameters were optimised for each of the nitro-explosives detected *via* direct infusion by auto-tuning for the $[M-H]^-$ or $[M+OAc]^-$ ions. Very little improvement was found for the other analytes when the RDX tune file was used as a starting point and so the same tune file was used across the run. As a result of this, there were no gaps in the run where non-target analytes could be missed due to changing acquisition files. A different tune file was required for the two peroxides which, unlike the nitro-explosives, were detected using positive ionisation, as $[M+H]^+$ or $[M+NH4]^+$ ions. Spray voltage, capillary temperature and HESI heater temperature were optimised by LC-HRMS analysis and the effect of each of these parameters on LC peak areas for the nitro-explosives are shown in Figure 2.12. Changing the spray voltage from -2.5 to -4.0 kV had very little effect on peak area, but temperature had a greater effect. The most temperature dependant analytes were the dinitrotoluenes (DNTs) which increased in intensity with temperature and RDX which decreased in intensity with increased temperature. Compromise temperatures of 300 °C for capillary temperature and 240 °C for heater temperature were selected.

Only twelve of the eighteen initial target analytes were detected using ESI and the ammonium acetate mobile phase, with ions for NB, the NT's and EGDN not detected. Detection of the other nitroesters was also problematic. Whilst the acetate adduct of NG, $[NG+OAc]^-$, has been detected in previous studies using an ESI source [44], here NG was only detected as a nitrate ion no matter what ionisation and ion optic parameters were used. This may be due to differences in the instrumentation used affecting adduct stability, but an exact cause was not found. As the nitrate

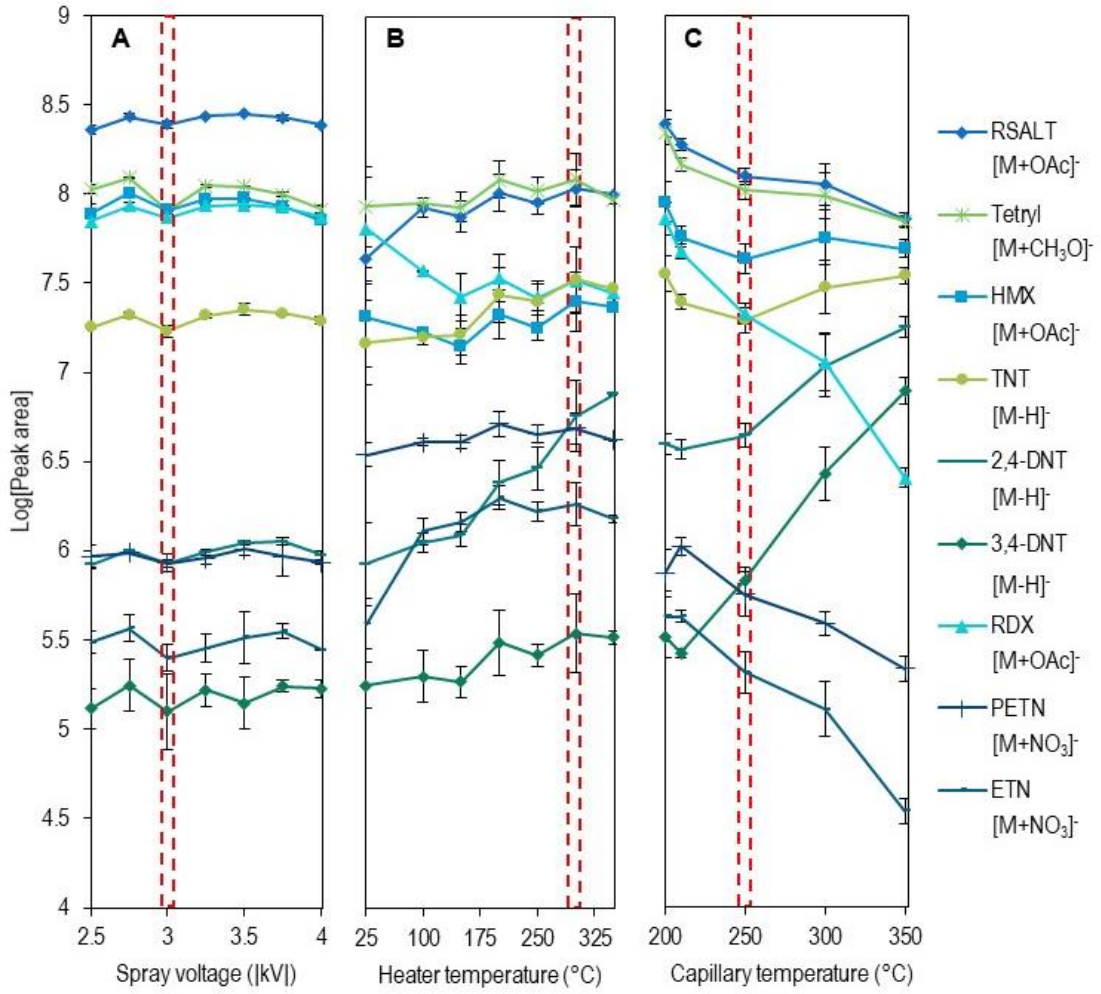


Figure 2.12: Optimisation of ESI-HRMS parameters for the nitro-explosives: a) spray voltage, b) HESI heater temperature and c) capillary temperature. Error bars represent standard deviation of n=3 repeats. Note. 2-NT, 3-NT, 4-NT, 2,6-DNT, NB, EGDN and NG were not detected as deprotonated molecules or adduct ions.

ion was also detected for ETN and PETN this was not deemed sufficient for identification and the tentative identification of NG was based on retention time rather than the detection of a unique ion. For forensic casework, further confirmation would be required. ETN was detected as the nitrate adduct, $[M+NO_3]^-$, but with low intensity and was not detected below 1.25 ng on column.

2.3.2.1.2 Effect of mobile phase additives on LC-(ESI)-HRMS

Due to the poor detection of the nitrate esters, nitrobenzene and nitrotoluenes with ammonium acetate in the mobile phase, alternative mobile phase additives were investigated. Previously Zhao *et al.* and Thomas *et al.* found, ammonium chloride aided detection of nitroesters [34,39] whilst Xu *et al.* found the use of no additives in the mobile phase best for nitroaromatics [40]. Therefore, the use of ammonium chloride and no mobile phase additives were both compared to the ammonium acetate mobile phase and a summary of the results is given in Table 2.4.

Table 2.4: LC-(ESI)-HRMS sensitivity and linear range for target nitro-explosives detected in negative mode using ammonium chloride, ammonium acetate or no mobile phase additive.

Analyte	Ammonium Chloride				Ammonium Acetate				No Additive			
	Ion	Cal 1 (pg/µL)	Linear range (pg/µL)	R^2	Ion	Cal 1 (pg/µL)	Linear range (pg/µL)	R^2	Ion	Cal 1 (pg/µL)	Linear range (pg/µL)	R^2
R-Salt	$[M+CHO_2]^-$	50	-	-	$[M+CHO_2]^-$	5	5 - 5000	0.992	$[M+CHO_2]^-$	5	5 - 5000	0.992
	$[M+Cl]^-$	5	5 - 2500	0.993	$[M+OAc]^-$	5	5 - 5000	0.994				
HMX	$[M+CHO_2]^-$	500	-	-	$[M+CHO_2]^-$	5	5 - 5000	0.992	$[M+CHO_2]^-$	5	5 - 5000	0.992
	$[M+Cl]^-$	5	5 - 1000	0.992	$[M+OAc]^-$	5	5 - 5000	0.994				
RDX	$[M+CHO_2]^-$	250	-	-	$[M+CHO_2]^-$	5	5 - 5000	0.993	$[M+CHO_2]^-$	5	50 - 5000	0.997
	$[M+Cl]^-$	5	5 - 2500	0.994	$[M+OAc]^-$	5	5 - 5000	0.994				
EGDN	$[M+Cl]^-$	5000	-	-	n.d.	-	-	-	n.d.	-	-	-
PA	$[M-H]^-$	5	5 - 5000	0.993	$[M-H]^-$	5	-	-	$[M-H]^-$	5	5 - 5000	0.994
NG	$[NO_3]^-$	25	25 - 5000	0.997	$[NO_3]^-$	25	25 - 5000	0.997	$[NO_3]^-$	25	25 - 5000	0.996
	$[M+NO_3]^-$	500	-	-	$[M+NO_3]^-$	n.d.	-	-	$[M+NO_3]^-$	25	25 - 5000	0.997
	$[M+Cl]^-$	10	10 - 5000	0.992	$[M+OAc]^-$	n.d.	-	-				
3,4-DNT	$[M-H]^-$	100	100-5000	0.994	$[M-H]^-$	2500	-	-	$[M-H]^-$	250	-	-
2,4/2,6-DNT*	$[M-H]^-$	25	25 - 5000	0.990	$[M-H]^-$	100	100 - 5000	0.999	$[M-H]^-$	50	50 - 5000	0.997
ETN	$[NO_3]^-$	25	25 - 2500	0.997	$[NO_3]^-$	25	25 - 5000	0.995	$[NO_3]^-$	50	-	-
	$[M+NO_3]^-$	100	500 - 5000	0.991	$[M+NO_3]^-$	250	-	-	$[M+NO_3]^-$	10	50 - 5000	0.991
	$[M+Cl]^-$	10	10 - 5000	0.998	$[M+OAc]^-$	n.d.	-	-				
Tetryl	$[M+CH_3O]^-$	50	100 - 1000	0.993	$[M+CH_3O]^-$	5	5 - 5000	0.992	$[M+CH_3O]^-$	5	50 - 2500	0.993
	$[M-NO]^-$	500	-	-	$[M-NO]^-$	5	5 - 5000	0.991	$[M-NO]^-$	5	-	-
	$[M+Cl]^-$	5	5 - 2500	0.995	$[M+OAc]^-$	n.d.						
TNT	$[M-H]^-$	5	5 - 5000	0.997	$[M-H]^-$	5	5 - 5000	0.996	$[M-H]^-$	5	5 - 2500	0.991
PETN	$[NO_3]^-$	5	5 - 5000	0.996	$[NO_3]^-$	5	5 - 5000	0.997	$[NO_3]^-$	10	50 - 5000	0.992
	$[M+NO_3]^-$	10	10 - 2500	0.993	$[M+NO_3]^-$	10	25 - 1000	0.997	$[M+NO_3]^-$	10	10 - 5000	0.997
	$[M+Cl]^-$	5	5 - 5000	0.994	$[M+OAc]^-$	250	-	-				

* Only one peak was detected in mixture for 2,4/2,6-DNT peak height was used for this peak, as it was not fully resolved from a peak of same accurate mass attributed to tetryl.
Cal 1- lowest concentration of standards run (5, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 2500 and 5000 pg μL^{-1}) in which analyte was detected, n.d. - not detected in 5000 pg μL^{-1}
Linear range - concentration range over which $R^2 > 0.99$ for bi-logarithmic linear trendline

Calibration curves were run with 8 mM ammonium acetate, 0.2 mM ammonium chloride or no additive in the mobile phase using a mixed standard at a range of concentrations from 5 – 5000 pg μL^{-1}. As the concentration used (5, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 2500 and 5000 pg μL^{-1}) covered multiple orders of magnitude, bi-logarithmic calibration lines were used to ensure a more even spread and therefore weighting of the data points. Sensitivity was compared based on the concentration of the lowest standard run in which the ion was detected, Calibrant 1 (Cal 1), since limits of detection were not explicitly determined yet. In many cases, where an ion was detected but no linear range is given, the bi-logarithmic calibration line fit well ($R^2 > 0.99$) to a quadratic equation instead.

As found previously [40], ionization of the nitroaromatics was most efficient with no mobile phase additives present. The ammonium chloride resulted in less ion suppression than the ammonium acetate, perhaps due to the lower concentration required (0.2 mM ammonium chloride compared to 8 mM ammonium acetate). Un-buffered mobile phase is generally not recommended because it led to less repeatable retention times, as shown by the size of error bars in Figure 2.13. The retention time of picric acid (PA) was affected the most by changing the mobile phase and the

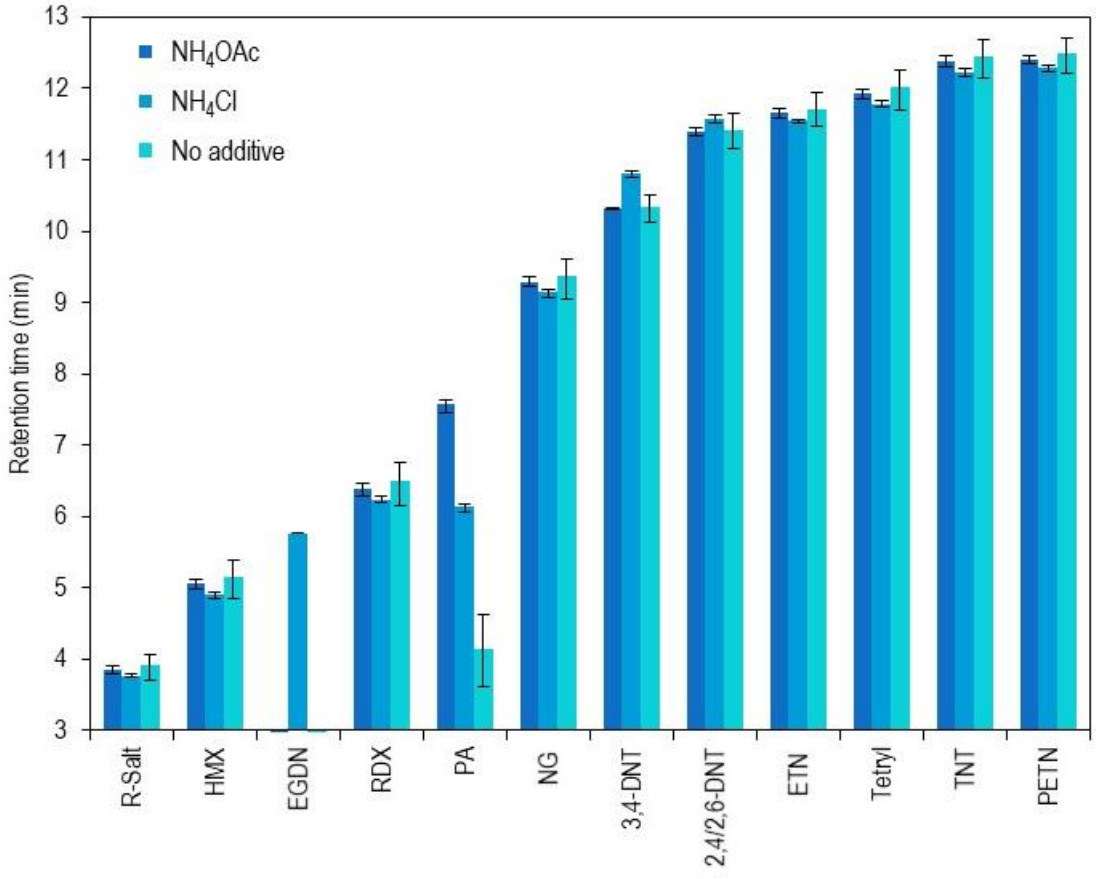


Figure 2.13: Average retention times for LC-(ESI)-HRMS with NH_4OAc, NH_4Cl or no additive. Error bars indicate maximum and minimum measured retention times, n=10.

peak shape was poor without any buffer. Picric acid was ionised with all mobile phases used (pH 6.4-7.8), due to the very acidic hydroxyl group attached to the benzene ring with three electron-withdrawing nitro groups (Figure 2.14).

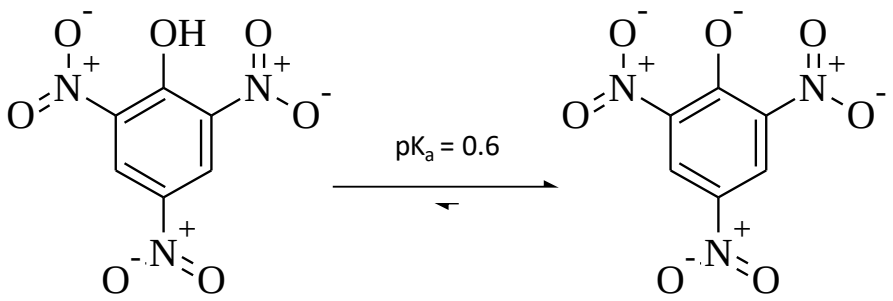


Figure 2.14: Chemical structure and acid dissociation constant (predicted using Percepta PhysChem Profiler) of picric acid, which was predominantly charged at the pH of all mobile phase.

In addition to sensitivity and linear range, it is also important to consider the type of ions detected when comparing mobile phases. Ions formed by deprotonation of the molecular ion or due to the addition of an adduct to the molecular ion were considered preferable and of greater forensic values than fragments such as nitrate which may be shared by multiple analytes. Chloride adducts were detected for all four nitrate esters with good sensitivity (except for EGDN) when ammonium chloride was present in the mobile phase. Therefore, despite the nitrate ion being detected with good sensitivity using all mobile phases, using ammonium chloride improved detection of the nitrate esters.

Overall, ammonium chloride was found to be the best mobile phase additive for this purpose and, since retention times varied somewhat, it was necessary to alter the gradient slightly in order to achieve good separation. The optimised separation with 0.2 mM ammonium chloride used a gradient of 40-90 % B over 15 min, followed by a wash step at 100% B for 5 minutes which was added after peaks were noticed in blanks peaks after prolonged use.

2.3.2.2 Optimisation of atmospheric pressure chemical ionisation – high resolution mass spectrometry parameters

Atmospheric pressure chemical ionisation mass spectrometry (APCI-MS) parameters were investigated and optimised for the initial 18 target analytes (Figure 2.15), using LC-MS with a 0.3 mL min^{-1} flow rate and 0.2 mM ammonium chloride in the mobile phases. All 18 target analytes were detected. Here, chromatographic peak height was used rather than peak area as two peaks were detected for 2,4- and 2,6-DNT but they were not quite resolved to baseline, preventing the use of peak area. Negative mode ionisation was used for all nitro-compounds and positive mode

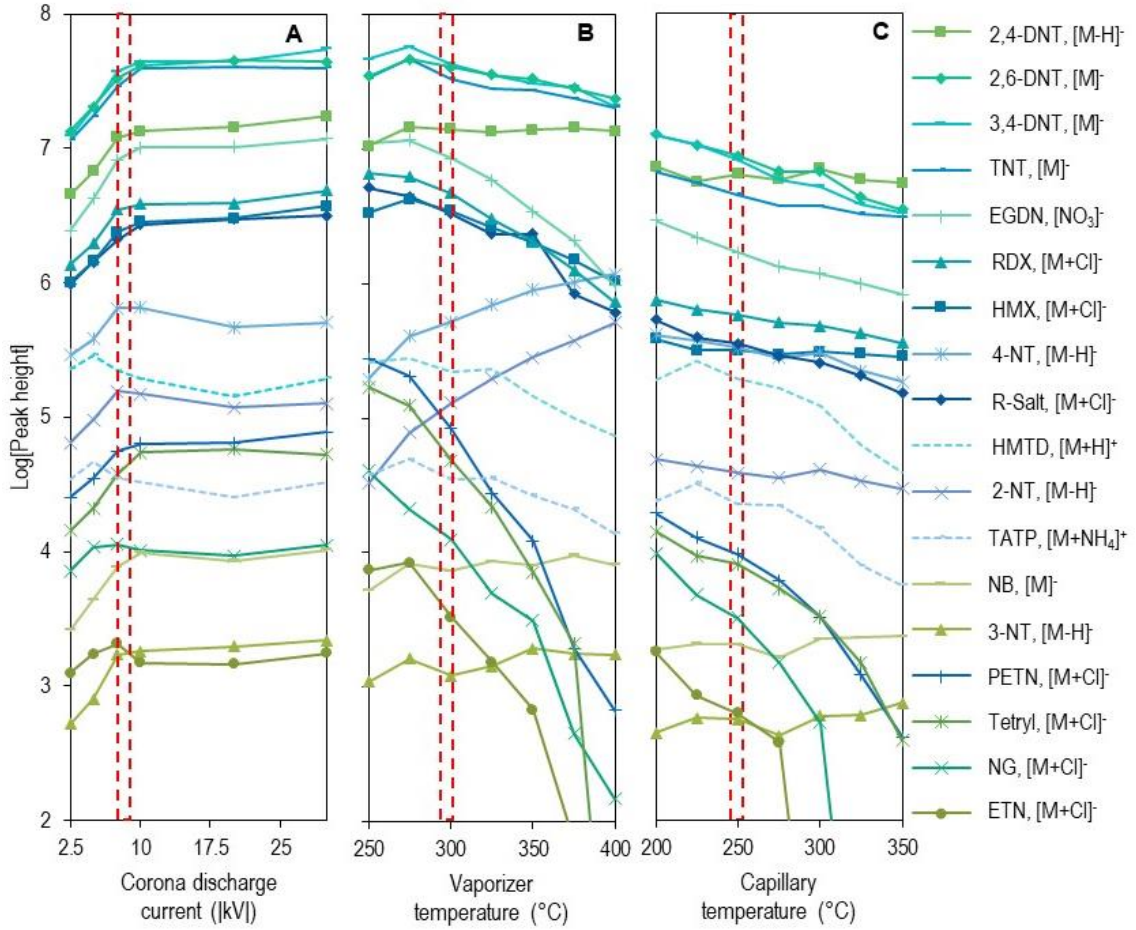


Figure 2.15: Optimisation of APCI parameters: a) corona discharge current, b) vaporiser temperature and c) capillary temperature. Dashed red boxes indicate the setting used for other optimisation experiments.

ionisation was used for the peroxides, TATP and HMTD. The dashed red boxes in Figure 2.15 indicate the settings used for the other two optimisation experiments. Therefore, in theory the three boxes should line up. Clearly this is not the case with chromatographic peak heights for the Capillary Temperature experiments appearing lower than for the other two experiments. This can be explained by the fact the Capillary Temperature experiments were run on a different date to the corona discharge current and vaporiser temperature when the overall sensitivity of the instrument was poorer, but the instrument still passed calibration.

Increasing corona discharge current led to an increase in chromatographic peak height until the effect levelled off after 10 kV which was selected as the corona discharge current (Figure 2.15 A). As was seen for ESI (Figure 2.12), the effect of APCI temperatures was compound dependent (Figure 2.15 B and C). With APCI, TNT and the dinitrotoluene isomers produced the strongest signal and were less effected by temperature than with ESI. However, two nitroaromatics (2-NT and 4-NT) still saw a marked increase in peak height with increased APCI vaporiser temperature.

As with ESI, increased temperatures had a detrimental effect on the peak height of the nitrate esters and to a lesser extent the nitramines. Therefore, a compromise was required when selecting vaporiser temperature and capillary temperature for a multi-analyte method. In this instance, a 300 °C vaporiser temperature and a 250 °C capillary temperature were selected for the optimised method. While these temperatures are lower than the typical 350 to 450 °C vaporiser temperature and 350 to 380 °C capillary temperature suggested by the manufacturer for this flow rate [101,102], they are not as low as the temperatures used in some methods for explosive analysis. For example, Xu et al. used a heated capillary temperature of 125 °C and vaporiser temperature of 160 °C [33]. In this study, one of the considerations when selecting the 250 °C capillary temperature was that lower capillary temperatures (200 °C and 225 °C) led to warnings about the vacuum and for the same reason temperatures below 200 °C were not investigated. In future work, it might be possible to overcome this challenge with a higher sheath gas setting than the 50 out of a maximum of 80 used here or a lower LC flow rate which would have resulted in a longer runtime. Xu et al. used a flow rate of 0.2 mL min^{-1} [33], which may explain the use of lower temperatures. DeTata et al. on the other hand used a slightly higher vaporiser temperature (325 °C) and a much higher flow rate (1 mL min^{-1}) [32].

2.3.2.3 Comparison of ESI and APCI for broad screening of organic explosives

Following optimization of ionization parameters for both electrospray ionization (ESI) (Figure 2.12) and atmospheric pressure chemical ionization (APCI) (Figure 2.15), a comparison of the two ionization techniques was performed for the initial target analytes (Figure 2.16). Both negative and positive ionization modes were required for detection of the 18 target compounds as the peroxides could only be detected in positive mode while the nitro-compounds were detected in negative mode. Molecular ions, (de)protonated molecules and chloride or ammonium adducts were used for all, except for EGDN where only an $[NO_3]^-$ fragment ion was detected.

Even at 25 ng on column, the three nitrotoluenes (2-NT, 3-NT and 4-NT), nitrobenzene (NB) and ethylene glycol dinitrate (EGDN) were not detected using ESI. Apart from EGDN, the nitrate esters formed chloride adducts and were detected with greater sensitivity by ESI than APCI. The nitramines (RDX and HMX) and nitrosamine (R-Salt) were also detected as chloride adducts but were detected with good sensitivity by both ionisation methods. The nitroaromatics on the other hand were detected with greater sensitivity by APCI, except for tetryl which could also be classed

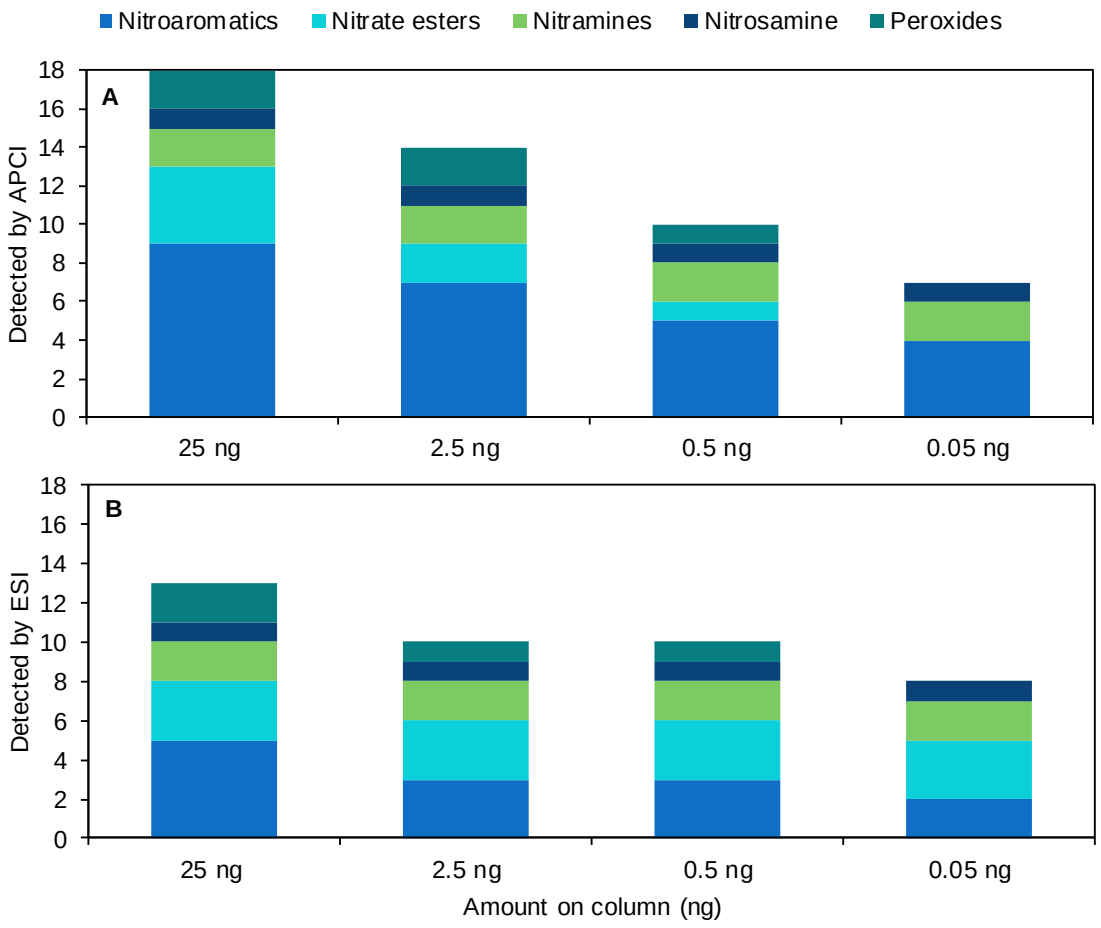


Figure 2.16: Number of initial target analytes detected by: A) APCI and B) ESI at 25, 2.5, 0.5 and 0.05 ng on column. N.b. for the purpose of this figure Tetryl is classed as a nitroaromatic not a nitramine.

as a nitramine and was the only nitroaromatic detected as a chloride adduct. All other nitroaromatics were detected as the molecular ion, $M^{-\cdot}$, and/or deprotonated molecule, $[M-H]^-$. The peroxides were also detected with greater sensitivity by APCI than ESI. Due to detection of all the target analytes, the APCI method was selected to test against a larger set of compounds for generalizability in Chapter 3.

2.3.3 Instrumental method performance of optimised multi-residue LC-APCI-HRMS method

A brief assessment of the optimized LC-APCI-HRMS method performance for the initial 18 target analytes is shown in Table 2.5. Method performance will be assessed in more detail and for a larger number of analytes in Chapter 3. Here, retention time ± standard deviation (SD), average *m/z* and mass accuracy are given for the molecular ion, (de)protonated molecule or adduct ion although these are not always the most intense ion. The only exception to this is EGDN where only the nitrate ion was detected. In Chapter 3, additional ions will also be considered for

screening and identification. In order to provide an indication of method sensitivity a calibration line was run and the lowest concentration in which the specified ion was detected (Cal 1) is given in Table 2.5. Determination of limits of detection for LC-HRMS methods is somewhat complicated by the lack of noise in many extract ion chromatograms (EICs). The method performance varied widely between analytes with TNT, 2,4-DNT, 2,6-DNT and 3,4-DNT molecular ions or deprotonated molecules detected in the lowest concentration standard analysed (1 pg μL^{-1}) but the ETN chloride adduct only detected in the highest concentration standard (10 ng μL^{-1}).

Table 2.5: APCI method performance for eighteen initial target analytes.

Analyte	t_R [a] (min)	SD (min)	Ion	Accurate *m/z* [b]	Mass accuracy (ppm)	Cal 1 [c] (pg μL^{-1})	Linear range [d] (pg)	Linearity (R^2)
HMTD	2.85	0.05	$[M+H]^+$	209.0763	-2.59	100	75-10000 [e]	0.998
R-Salt	3.6	0.03	$[M+Cl]^-$	209.0194	-3.44	2.5	2.5-1000	0.995
HMX	4.7	0.07	$[M+Cl]^-$	331.0157	-2.31	2.5	2.5-1000	0.997
EGDN	5.5	0.04	$[NO_3]^-$	61.9884	1.14	250	750-10000	0.995
RDX	6.1	0.07	$[M+Cl]^-$	257.0041	-2.82	2.5	2.5-1000	0.997
NB	7.8	0.08	$[M]^{-\cdot}$	123.0326	-0.06	1000	2500-10000	0.97
NG	9.2	0.08	$[M+Cl]^-$	261.9719	-2.39	250	250-10000 [e]	0.983
3,4-DNT	10.4	0.1	$[M]^{-\cdot}$	182.0332	-0.67	1	1-1000	0.993
2-NT	10.7	0.1	$[M-H]^-$	136.0403	-0.58	250	250-1000	0.984
4-NT	11	0.13	$[M-H]^-$	136.0403	-0.52	500	500-1000	0.999
3-NT	11.1	0.01	$[M]^{-\cdot}$	137.0482	-0.49	7500	-	-
2,6-DNT	11.3	0.11	$[M]^{-\cdot}$	182.0331	-0.87	1	1-1000	0.993
2,4-DNT	11.4	0.1	$[M-H]^-$	181.0253	-0.87	1	1-1000	0.998
ETN	12	0.09	$[M+Cl]^-$	336.9681	1.35	10000	-	-
Tetryl	12.1	0.12	$[M+Cl]^-$	321.9831	-2.07	100	100-10000 [e]	0.996
TATP	12.22	0.08	$[M+NH_4]^+$	240.1441	-0.46	250	250-10000	0.998
TNT	12.5	0.11	$[M]^{-\cdot}$	227.0182	-0.81	1	1-1000	0.993
PETN	12.7	0.09	$[M+Cl]^-$	350.9832	-1.83	75	75-10000 [e]	0.994

[a] Average of n=18
[b] Average of n=6
[c] Lowest concentration of standards run (1, 2.5, 5, 10, 7.5, 25, 50, 75, 100, 250, 500, 750, 1000, 2500, 5000, 7500 and 10000 pg μL^{-1}) in which analyte was detected
[d] Linear range for bi-logarithmic calibration line
[e] Fitted to a non-linear trendline.

2.3.4 Application of LC-HRMS method to the detection of explosives in contact traces

2.3.4.1 Recovery of organic explosives from latent fingermarks

Recovery of organic explosives from latent fingermaks was determined by spiking six fingermarks with a mixed standard containing the initial 16 nitro-explosives plus nitromethane (NM), picric acid (PA), diethylene glycol dinitrate (DEGDN) and trinitrobenzene (TNB). The results of this experiment are shown in Figure 2.17. Due to the variability of fingermarks even from the same donor, it was not possible to produce a matrix matched standard and so the recoveries given are

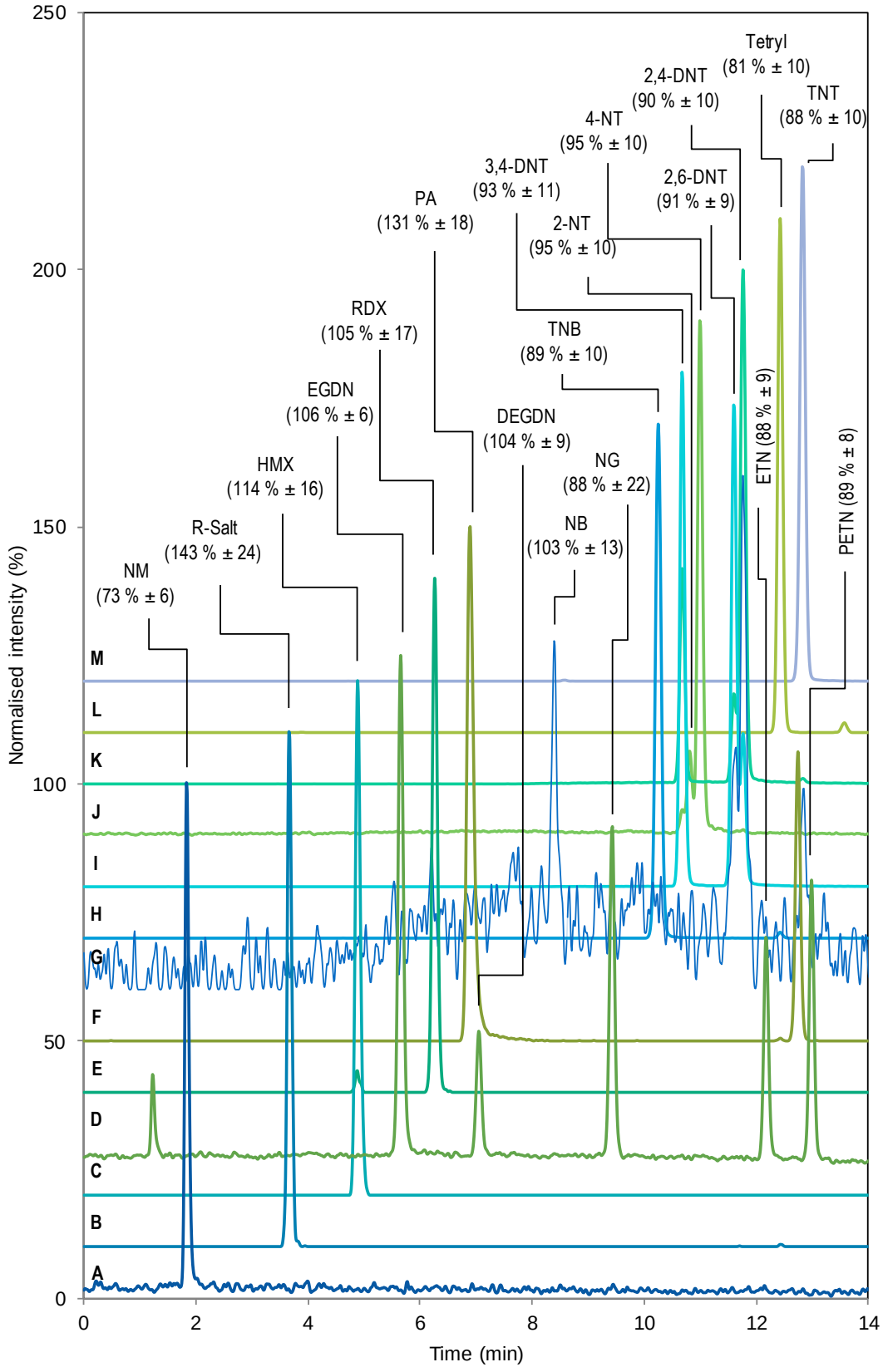


Figure 2.17: Example of spiked fingermark with average recoveries given in brackets ± standard deviation (n=6). LC-(APCI)_HRMS extracted ion chromatograms with a 5 ppm mass accuracy for A: *m/z* 60.00910, B: *m/z* 209.01955, C: *m/z* 331.01593, D: *m/z* 61.98837, E: *m/z* 257.00430, F: *m/z* 227.98983, G: *m/z* 123.03258, H: *m/z* 213.00274, I: *m/z* 182.03331, J: *m/z* 136.04040, K: *m/z* 181.02548, L: *m/z* 241.02146 and M: *m/z* 227.01839.

relative to a standard in methanol. As a result of this, stated recoveries, which were all > 70 %, may have included enhancement or suppression due to the matrix as well as the percentage recovery of the extraction method.

Of the 20 explosives analyzed, 19 were detected with recoveries greater than 70 %. 3-NT was not detected at all which was not surprising given that even 100 % recovery would have resulted in a concentration (1 ng μL^{-1}) lower than the instrumental LOD for 3-NT. Instrumental sensitivity was also a challenge for NB which was detected close to LOD, with large background noise relative to signal obtained. The nitrate esters were only detected as nitrate ions but with good calculated recoveries (88-106 %). Based on the assessment of instrumental method performance shown in Table 2.5, the chloride adducts for EGDN and ETN were expected to be below the LOD.

The overlaid extracted ion chromatograms (EICs) shown in Figure 2.17 also demonstrate the chromatographic separation achieved by the optimised LC-HRMS method. Each EIC was normalized to 100 % but the target analyte was not always the only peak present. As expected the nitrate ion was not very selective with peaks present for each of the nitrate esters included in the mix. All three dinitrotoluenes produce both the $[M]^-$ and $[M-H]^-$ ions and while 2,4-DNT and 2,6-DNT were not fully baseline resolved some separation was achieved and they could also be distinguished based on the fact 2,4-DNT forms more of the deprotonated molecule while 2,6-DNT forms more of the molecular ion. In the case of NB, the target analyte was not responsible for base peak as there was a larger peak corresponding to the retention time of 2,4-DNT in the EIC for *m/z* 123.03258 ± 5 ppm. Additionally, there was a second peak in the RDX chloride adduct EIC (*m/z* 257.00430 ± 5 ppm) corresponding to HMX retention time. The effect of common ions on selectivity will be considered in more detail in Chapter 3.

2.3.4.2 Detection of trace explosives in latent human fingermarks after handling commercial explosives

In order to test the method under more realistic conditions, a fingermark depletion series was carried out after the donor had handled a bulk explosive. The results for the 1st, 11th, 21st, 31st, 41st, 51st and 61st marks are shown in Figure 2.18, along with a blank fingermark (from a donor who had not been in contact with explosives) and 0.01 ng μL^{-1} standards for comparison. A peak for the RDX chloride adduct was visible in all seven depletion series samples. A change in signal was also observed for the HMX chloride adduct, in all but the 51st fingermark sample, but at a much lower intensity. Without any background noise to determine a 3:1 signal:noise ratio for the

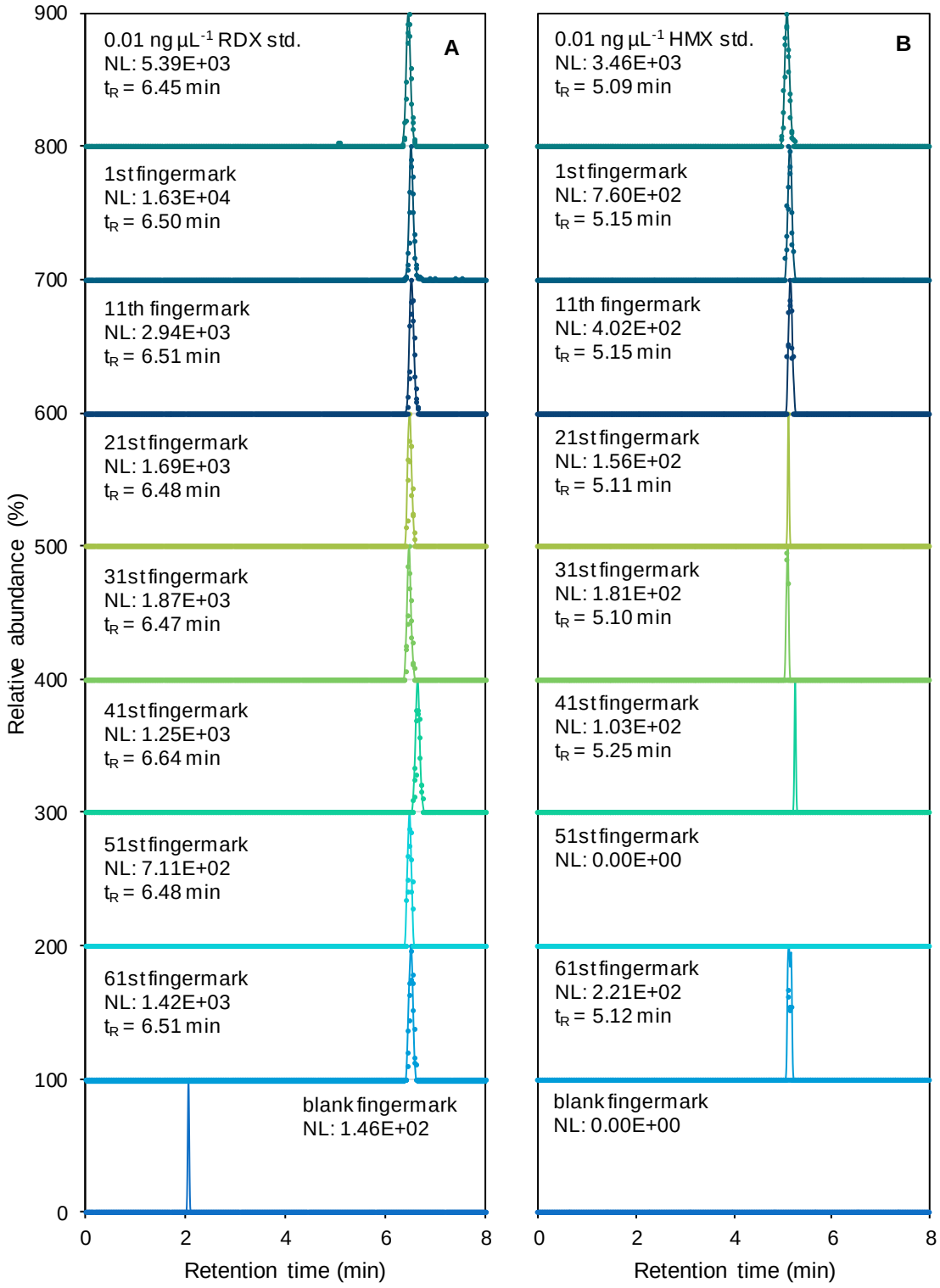


Figure 2.18: Fingermark depletion series following contact with a commercial explosive. Part A shows overlaid LC-(APCI)-HRMS extracted ion chromatograms for $[RDX+Cl]^-$, *m/z* 257.00430 ± 5 ppm and part B shows overlaid extracted ion chromatogram for $[HMX+Cl]^-$, *m/z* 331.01593 ± 5 ppm. Data points represent unsmoothed data and lines are plotted with 7 point gaussian smoothing. NL = normalised intensity represented by 100%.

limit of detection it was difficult to assess the significance of this change in signal and whether or not it can actually be considered a peak. Table 2.6 shows which samples contained a peak detected by Xcalibur Qual Browser software using the default peak detection settings, with and without 7-point gaussian smoothing. Whether or not smoothing was used affected the results with peaks detected for RDX in all the smoothed depletion series samples but not in the 41st and 61st samples without smoothing. For HMX peaks were only detected in the 1st and 11th fingermarks with smoothing and, more surprisingly, the 61st fingermark without smoothing. Additional detection criteria, such as a peak height threshold, would change the results again, with no RDX peak detected in the 51st fingermark and no HMX peaks detected in any of the depletion series samples above 1000 counts.

Table 2.6: Compounds detected by LC-(APCI)-HRMS in fingermark depletion series using default peak detection settings with and without smoothing.

	RDX (*m/z* 257.00430 ± 5 ppm)		HMX (*m/z* 331.01593 ± 5 ppm)	
Sample	Smoothed*	Unsmoothed	Smoothed*	Unsmoothed
0.1 ng µL^{-1}	PEAK	PEAK	PEAK	PEAK
0.01 ng µL^{-1}	PEAK	PEAK	PEAK	PEAK
1st Fingermark	PEAK	PEAK	PEAK	NO PEAK
11th Fingermark	PEAK	PEAK	PEAK	NO PEAK
21st Fingermark	PEAK	PEAK	NO PEAK	NO PEAK
31st Fingermark	PEAK	PEAK	NO PEAK	NO PEAK
41st Fingermark	PEAK	NO PEAK	NO PEAK	NO PEAK
51st Fingermark	PEAK	PEAK	NO PEAK	NO PEAK
61st Fingermark	PEAK	NO PEAK	NO PEAK	PEAK
Blank Fingermark	NO PEAK	NO PEAK	NO PEAK	NO PEAK

*7-point gaussian smoothing performed in Xcalibur Qual browser

The ambiguity regarding LC-HRMS peak detection when approaching the detection limits of the instrument is problematic, particularly for the use of LC-HRMS in forensic science where results need to stand up in a court of law. For confident detection and identification of explosives the sensitivity and selectivity of the method requires further investigation and will be revisited in Chapter 3.

2.4 Conclusion

An LC-HRMS method was developed for 18 initial target analytes, including 9 nitroaromatics, 4 nitrate esters, 3 nitramines, 1 nitrosamine and 2 peroxides. Separation of 20 forensically relevant explosives, including 16 of the initial target analytes was achieved by HPLC-UV using an ACE C_{18} Ar column. All analytes were well separated and had good resolution, except for isomers 2,4- and 2,6-DNT which only had a resolution of 0.87. Ionisation of the 18 initial target analytes was achieved using APCI, in negative mode for the nitro explosives and positive mode for the peroxides. Ammonium chloride was used in the mobile phase to facilitate formation of chloride adducts in negative mode and ammonium adducts in positive mode. Compromise conditions were required to allow detection of all target analytes and, by using these conditions, the main cost was in the sensitivity of the nitrate ester detection. Using ESI and lower ionisation temperatures would be preferable if a targeted method just for nitrate esters was required.

In optimisation experiments, the optimised conditions will inevitably be influenced by the starting conditions. Here, the majority of the chromatographic optimization was performed using LC-UV not LC-HRMS and with ammonium acetate in the mobile phase, not ammonium chloride. Changing to ammonium chloride had some effect on the separation and so the results of the column selection and gradient optimisation experiments may have differed if performed using an ammonium chloride mobile phase. In hindsight, it may have been better if ionisation conditions and in particular the use of mobile phase additives was optimised first. However, at the beginning of this project it was assumed that ammonium acetate would be a suitable mobile phase additive, since it had been used to aid ionisation in previous methods. Changing from UV to HRMS detection also affected the chromatographic separation as there was a bigger difference in instrumental response between analytes when using LC-HRMS. This meant that while two peaks with a similar height by UV may be resolved, if one was much larger than the other by HRMS it could become a shoulder and not fully resolved. This could be an argument against the use of HPLC-UV for the optimisation of HPLC-HRMS methods. However, the cost of instrument hire meant it was not feasible to perform all optimisation experiments by LC-HRMS for this project. Similarly, the comparison of mobile phase additives was performed using ESI rather than APCI which might perform differently with different mobile phases.

The presented optimized method may not be the optimum method for explosives detection, but it was an improvement upon the existing method that was used as a starting point. Further

improvements may be seen if optimization experiments were repeated using the optimized conditions as a starting point. However, provided a method is fit for purpose there is little value in continuing optimization indefinitely. Additionally, optimum conditions are likely to vary between instruments with differences in the architecture of the ion source and subsequent ion transfer, as was seen in this study by the lack of detection of the nitroglycerin acetate adduct using the Exactive™ mass spectrometer.

The method was applied to the detection of explosives in contact traces, with 19 nitro explosives detected in spiked fingermarks. RDX and HMX chloride adduct, extracted ion chromatograms for a fingermark depletion series were also presented. While RDX was clearly detected following handling of a commercial explosive, these results raised questions regarding detection criteria when using LC-HRMS. For example, without any background noise, what change in signal is required for a peak to be detected, how many consecutive scans does this signal change need to be detected in and is smoothing beneficial or misleading?

In this chapter, full-scan HRMS has not yet been used to its full potential and the method performance of only one ion (typically the molecular ion, (de)protonated molecule or adduct ion) has been considered for each analyte. In the next chapter, the method performance including both selectivity and sensitivity will be considered in greater detail along with effect of using multiple ions, fragments and isotope ratios for detection and identification. Additionally, the method performance of compounds not considered during the optimization phase will also be investigated to assess non-target method performance and the use of HRMS for suspect screening.

Chapter 3: Performance of LC-HRMS method against an expanded set of compounds

3.1 Introduction

High resolution mass spectrometry coupled to liquid chromatography (LC-HRMS) has become a popular choice for screening and confirmation of explosives due to the ability to detect numerous compounds with high mass resolution and high mass accuracy in a single method [32,33]. As discussed in Chapter 2, the use of full-scan HRMS allows more explosives to be included in a method, without sacrificing the number of data points per chromatographic peak, since data for all ions is collected in each scan. Another advantage with full-scan HRMS is that since data is collected for all ions, new compounds may also be detected through suspect screening, providing they are sufficiently ionised. Post-acquisition data mining has already been used in this way for a range of applications including: drugs and pharmaceuticals in the environment [103], contaminants and additives in food [104,105] and new psychoactive substances [106] by LC-HRMS. Non-target analysis can be particularly challenging for explosive compounds, which unlike most drugs, tend not to be very acidic or basic and are therefore difficult to ionise. Additionally, it is not possible to optimise and performance test a method for individual non-target compounds if it is not known in advance that they will be of interest. Therefore, in Chapter 2 an LC-HRMS method was optimised for 18 representative explosives with a range of functional groups to maximise applicability to new compounds. The first aim of this chapter is to assess the LC-HRMS method performance for 62 compounds, including an additional 44 compounds for which the method has not been optimised, to mimic the situation with suspect and biased non-target screening. Qualitative method performance was assessed only, since selectivity (both between explosives and from matrix) and limits of detection and/or identification are most important for screening and identification purposes.

Greater confidence of identification is possible using HRMS as it can separate isobaric ions with the same nominal mass but different exact masses. However, questions remain regarding how much greater this confidence is, how it can be measured and what this means for identification criteria for explosives using HRMS methods. Xu et al. adapted the identification point system initially defined in the European Commission decision (2002/657/EC) to allocate additional points to orbitrap MS ions [33,93]. According to the original identification points system, two identification points (IPs) could be earned per HRMS ion. Therefore, at least 2 ions were required to achieve the 4 IPs needed for confirmation of 'substances having anabolic effect and unauthorized substances' or the 3 IPs needed for confirmation of veterinary drugs and contaminants. A

minimum of at least one ion ratio within the permitted tolerance (20-50 % depending on the relative intensity) was also required in both cases [93]. Xu et al. argued that the greater mass resolution achieved by orbitrap MS (≥ 60,000 FWHM) compared to the definition of HRMS used in the European Commission decision (~20,000 FWHM) warranted 3 IPs per orbitrap HRMS ion, with an additional IP available when the isotopic pattern and isobaric profile fit to the theoretical one [33]. Based on these assignments the required 3 or 4 IPs could be reached with a single orbitrap HRMS ion. Xu et al. also interpreted the ion ratio requirement as only necessary if calculating the sum of IPs for different ions, rather than essential for confirmation. Allowing a single HRMS ion to provide enough IPs for identification, ignores potential problems with common ions and isomers with the same elemental composition. Also, increased resolution may not automatically result in increased confidence, especially if the mass accuracy remains the same. Increased resolution may physically separate an analyte from more interferences but without increased mass accuracy it would still be unclear whether a peak was due to the analyte or interferences. The main increase in confidence provided by higher mass resolution, when combined with improved mass accuracy, is fewer possible elemental compositions. If the achievable mass accuracy remains unchanged then, even with increased resolution, there would be no change in the number of possible elemental compositions. Additionally, there will be a limit to the additional value and confidence gained by increased resolution and accuracy, as once the elemental composition of an ion has been unequivocally identified, no additional information can be gained by either increased resolution or through matching isotopic patterns.

Lehotay et al. argued that “the identification-point system is not scientific” with the number of IPs required for identification arbitrary, rather than based on a rigorous assessment of the number of false positives and negatives [94], and the same could be said for increasing the number of IPs awarded to an HRMS ion to 3. Lehotay et al. also pointed out that HRMS does not always provide a unique elemental composition, especially without any elements with large mass defects, and HRMS cannot distinguish between isomers with the same elemental composition [94]. Therefore, there is a limit to how much additional selectivity can be provided by increased mass resolution, mass accuracy and isotopic patterns. These factors can only confirm the elemental composition of the ion, which is not the same as the identity of the compound.

Mol et al. found relying on the exact mass and retention time of a single diagnostic ion for qualitative screening of pesticides resulted in too many false positives and recommended the use

of isotopes or fragments as secondary diagnostic ions [104]. When using less common isotopes as a second diagnostic ion, Mol et al. reduced the number of false positives from 600 to 36 [104]. Whilst consideration of isotopes is clearly of benefit, 36 false positives are still too many for confirmatory identification, especially for forensic purposes where they could contribute towards a miscarriage of justice. Also, isotopes will only ever be able to confirm the elemental composition and, unlike fragment ions, do not provide any additional structural information. A more recent guidance document, on analytical quality control and method validation procedures for pesticide residues and analysis in food and feed, by the European Commission recommends the use of at least 2 ions with a mass accuracy of ≤ 5 ppm (< 1 mDa for m/z < 200), preferably including the molecular ion, (de)protonated molecule or adduct ion and including at least one fragment ion [107]. Again, evidence to support the use of these identification requirements for the forensic identification of explosives is currently lacking. Use of stricter identification requirements may lead to more false negatives and for screening methods, minimising false negatives is of greater importance than minimising false positives [93].

Therefore, the second aim of this chapter was to investigate the effect of different identification criteria on selectivity and limits of detection to further understand identification requirements for explosive compounds by LC-HRMS. Theoretical experiments were performed to determine the number of elemental compositions within 5 ppm of the exact mass and the number of isomers with the same elemental composition. The effect of mass accuracy, number and type of ions and retention time windows on selectivity and detection limits was then investigated. This will provide evidence to support the selection of evidence-based identification requirements in the future.

3.2 Experimental

3.2.1 Materials

Reference materials for 62 explosives, precursors or transformation products were either purchased from Kinesis Ltd. (St Neots, UK) or Sigma-Aldrich (Gillingham, UK), or donated by the Forensic Explosives Laboratory (FEL). Details of the compounds used can be found in Table 3.1. For most compounds (n=53), purity was greater than 98 %. No purity information was available for diacetone diperoxide (DADP), mannitol hexanitrate (MHN), methyl ethyl ketone peroxide (MEKP) or sorbitol hexanitrate (SHN) which were donated by FEL. Therefore, stated concentrations for these compounds were over estimations based on 100 % purity. HPLC grade methanol (Fisher Scientific UK, Loughborough, UK) and ultrapure water (18.2 MΩ.cm) delivered from a Millipore Milli-Q water ultra-purification system (Millipore, Bedford, MA, USA) were used for mobile phases along with ammonium chloride which was purchased from Sigma-Aldrich. All mobile phases were sonicated for 15 min and filtered before use.

3.2.2 Liquid chromatography-high resolution mass spectrometry conditions

The LC-HRMS method optimised during Chapter 2, with an ACE C_{18} Ar column and atmospheric pressure chemical ionisation (APCI) in either positive or negative mode, was used for all analysis. For experiments investigating the use of higher-energy C-trap dissociation (HCD), a method alternating between HCD with a collision energy of 20 and full-scan MS was compared to the full-scan MS method developed during Chapter 2.

3.2.3 Assessing instrumental method performance

Calibration lines were run for 49 compounds in a mixed standard. Unretained compounds (e.g. aspirin, erythritol, mannitol, salicylic acid, sorbitol, urea and xylitol), compounds where purity was unknown (e.g. MHN, SHN, DADP and MEKP) and compounds that could not be distinguished from other compounds in the mixed standard (e.g. 1-MNG and ammonium picrate) were not included. All other compounds with a signal/noise ≥ 5 x 10^5 at 10 ng μL^{-1} were included at 1, 2.5, 5, 7.5, 10, 25, 50, 75, 100, 250, 500, 750 and 1000 pg μL^{-1}. Compounds with a signal/noise ≤ 5 x 10^5 at 10 ng μL^{-1} were included at 10 times these concentrations. This was due to significant differences in response between compounds, with the most sensitive compounds plateauing at the higher concentrations, but the least sensitive compounds detected at insufficient concentrations to plot a calibration line. Six replicate injections were also analysed at three different concentrations (5, 50 and 500 pg μL^{-1} or 10 times these concentrations) on the same

Table 3.1: Purity, concentration and source of reference materials.

Analyte	Abbreviation	IUPAC name	Purity (%)	Solvent	Concentration (µg mL^{-1})	Source
1,2-Diaminopropane	1,2-DAP	propane-1,2-diamine	99.8	CH_3OH	100	Kinesis ltd.
1,2-Dinitrobenzene	1,2-DNB	1,2-Dinitrobenzene	≥ 99	CH_3OH	1000	Kinesis ltd.
1,2-Dinitroglycerin	1,2-DNG	3-Hydroxy-1,2-propanediyl dinitrate	98.6	CH_3CN:CH_3OH	100	Kinesis ltd.
1,3-Dinitrobenzene	1,3-DNB	1,3-Dinitrobenzene	97	CH_3CN:CH_3OH	1000	Kinesis ltd.
1,3-Dinitroglycerin	1,3-DNG	2-Hydroxy-1,3-propanediyl dinitrate	99.3	CH_3CN:CH_3OH	100	Kinesis ltd.
1-Nitroglycerin	1-MNG	2,3-Dihydroxypropyl nitrate	99.8	CH_3CN:CH_3OH	100	Kinesis ltd.
2,4-Diamino-6-nitrotoluene	2,4-DA-6-NT	4-Methyl-5-nitro-1,3-benzenediamine	99	CH_3CN	100	Kinesis ltd.
2,4-Dinitrotoluene[a]	2,4-DNT	1-Methyl-2,4-dinitrobenzene	≥ 99.9	CH_3OH	1000	Kinesis ltd.
2,6-Bis(Picrylamino)-3,5-dinitropyridine	PYX	3,5-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-2,6-pyridinediamine	98.3	CH_3CN	100	Kinesis ltd.
2,6-Diamino-4-nitrotoluene	2,6-DA-4-NT	2-Methyl-5-nitro-1,3-benzenediamine	99.7	CH_3CN	100	Kinesis ltd.
2,6-Dinitrotoluene[a]	2,6-DNT	2-methyl-1,3-dinitrobenzene	≥ 99.9	CH_3OH	1000	Kinesis ltd.
2-Amino-4,6-dinitrotoluene	2-A-4,6-DNT	2-Methyl-3,5-dinitroaniline	100	CH_3CN:CH_3OH	1000	Kinesis ltd.
2-Nitroglycerin	2-MNG	1,3-Dihydroxy-2-propanyl nitrate	99	CH_3CN:CH_3OH	100	Kinesis ltd.
2-Nitrotoluene[a]	2-NT	1-Methyl-2-nitrobenzene	99.0	CH_3CN:CH_3OH	1000	Kinesis ltd.
3,4-Dinitrotoluene[a]	3,4-DNT	4-methyl-1,2-dinitrobenzene	100.0	CH_3OH	1000	Kinesis ltd.
3,5-Dinitroaniline	3,5-DNA	3,5-Dinitroaniline	100	CH_3CN:CH_3OH	100	Kinesis ltd.
3-Nitrotoluene[a]	3-NT	1-Methyl-3-nitrobenzene	98.7	CH_3CN:CH_3OH	1000	Kinesis ltd.
4-Amino-2,6-dinitrotoluene	4-A-2,6-DNT	4-Methyl-3,5-dinitroaniline	100	CH_3CN:CH_3OH	1000	Kinesis ltd.
4-Nitrotoluene[a]	4-NT	1-Methyl-4-nitrobenzene	99.2	CH_3CN:CH_3OH	1000	Kinesis ltd.
Acetyl salicylic acid	Aspirin	2-Acetoxybenzoic acid	≥ 99	Solid	-	Sigma-Aldrich
Ammonium picrate		Ammonium 2,4,6-trinitrophenolate	99.8	CH_3CN	100	Kinesis ltd.
Diacetone diperoxide	DADP	3,3,6,6-Tetramethyl-1,2,4,5-tetroxane	N/A	CH_3CN	~1000	FEL
Diazodinitrophenol	DDNP	6-Diazonio-2,4-dinitro-2,4-cyclohexadien-1-olate	99.5	CH_3CN	1000	Kinesis ltd.

Table 3.1 (Continued): Purity, concentration and source of reference materials.

Analyte	Abbreviation	IUPAC name	Purity (%)	Solvent	Concentration (μg mL^{-1})	Source
Diethyl diphenylurea	DEDPU	1,3-Diethyl-1,3-diphenylurea	99	Solid	-	Sigma-Aldrich
Diethylene glycol dinitrate	DEGDN	Oxydi-2,1-ethanediyl dinitrate	99.9	CH_3CN:CH_3OH	100	Kinesis ltd.
Dimethyl diphenylurea	DMDPU	1,3-Dimethyl-1,3-diphenylurea	99	Solid	-	Sigma-Aldrich
Dimethyldinitrobutane	DMNB	2,3-Dimethyl-2,3-dinitrobutane	98	Solid	-	Sigma-Aldrich
Dinitrosalicylic acid	DNSA	2-Hydroxy-3,5-dinitrobenzoic acid	98	Solid	-	Sigma-Aldrich
Diphenylamine	DPA	N-Phenylaniline	≥ 99	CH_3CH_2OH	1000	Kinesis ltd.
Erythritol		(2R,3S)-1,2,3,4-Butanetetrol	≥ 99	Solid	-	Sigma-Aldrich
Erythritol tetranitrate[a]	ETN	(2S,3R)-3,4-Bis(nitrooxy)-1,2-butanediyl dinitrate	99.9	CH_3CN	1000	Kinesis ltd.
Ethylene glycol dinitrate[a]	EGDN	1,2-Ethanediyl dinitrate	96.2	CH_3CN	100	Kinesis ltd.
Hexahydro-1,3,5-trinitro-1,3,5-triazine[a]	RDX	1,3,5-Trinitro-1,3,5-triazinane	98.6	CH_3CN:CH_3OH	1000	Kinesis ltd.
Hexahydro-1,3,5-trinitroso-1,3,5-triazine[a]	R-Salt	1,3,5-Trinitroso-1,3,5-triazinane	99.8	CH_3CN	100	Kinesis ltd.
Hexamethylene diperoxide diamine[a]	HMDD	3,4,8,9-Tetraoxa-1,6-diazabicyclo[4.4.2]dodecane	≥ 99	CH_3CN	100	Kinesis ltd.
Hexamethylene triperoxide diamine[b]	HMTD	3,4,8,9,12,13-Hexaoxa-1,6-diazabicyclo[4.4.4]tetradecane	98.4	CH_3CN	100	Kinesis ltd.
Hexamethylenetetramine	HMTA	1,3,5,7-Tetraazatricyclo[3.3.1.$1^{3,7}$]decane	≥ 99	Solid	-	Sigma-Aldrich
Hexanitrodiphenylamine	HNDPA	2,4,6-Trinitro-N-(2,4,6-trinitrophenyl)aniline	97.9	CH_3CN:CH_3OH	100	Kinesis ltd.
Mannitol		D-Mannitol	≥ 98	Solid	-	Sigma-Aldrich
Mannitol hexanitrate	MHN	1,2,3,4,5,6-Hexa-O-nitro-D-mannitol	N/A	CH_3CN	~1000	FEL
Methyl ethyl ketone peroxide[c]	MEKP	*Depends on oligomer*	N/A	CH_3OH	~100	FEL
Nitrobenzene[a]	NB	Nitrobenzene	99.8	CH_3OH	1000	Kinesis ltd.
Nitroglycerin[a]	NG	1,2,3-Propanetriyl trinitrate	99.4	97CH_3CH_2OH:3CH_3OH	1000	Kinesis ltd.
Nitroguanidine	NQ	1-Nitroguanidine	≥ 99	CH_3OH	100	Kinesis ltd.
Nitromethane	NM	Nitromethane	≥ 99	CH_3OH	100	Kinesis ltd.
N-methyl-N,2,4,6-tetranitroaniline[a]	Tetryl	N-Methyl-N,2,4,6-tetranitroaniline	98.6	CH_3CN:CH_3OH	1000	Kinesis ltd.
Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazine[a]	HMX	1,3,5,7-Tetranitro-1,3,5,7-tetrazocane	98.0	CH_3CN:CH_3OH	1000	Kinesis ltd.
Pentaerythritol tetranitrate[a]	PETN	3-(Nitrooxy)-2,2-bis[(nitrooxy)methyl]propyl nitrate	99.3	CH_3OH	1000	Kinesis ltd.

Table 3.1 (Continued): Purity, concentration and source of reference materials.

Analyte	Abbreviation	IUPAC name	Purity (%)	Solvent	Concentration (µg mL^{-1})	Source
Picramic acid		2-Amino-4,6-dinitrophenol	≥ 99	CH_3CN:CH_3OH	100	Kinesis ltd.
Picric acid	PA	2,4,6-Trinitrophenol	99.1	CH_3CN:CH_3OH	100	Kinesis ltd.
Propylene glycol dinitrate	PGDN	1,2-Propanediyl dinitrate	99.4	CH_3OH	100	Kinesis ltd.
Salicylic acid		Salicylic acid	≥ 99	Solid	-	Sigma-Aldrich
Sorbitol		D-Glucitol	≥ 98	Solid	-	Sigma-Aldrich
Sorbitol hexanitrate	SHN	1,2,3,4,5,6-Hexa-O-nitro-D-glucitol	N/A	CH_3CN	~1500	FEL
Triacetone triperoxide[a]	TATP	3,3,6,6,9,9-Hexamethyl-1,2,4,5,7,8-hexoxonane	99.9	CH_3CN	100	Kinesis ltd.
Triaminotrinitrobenzene	TATB	2,4,6-Trinitro-1,3,5-benzenetriamine	99.6	$(CH_3)_2NCH$	40	Kinesis ltd.
Triethylene glycol dinitrate	TEGDN	1,2-Ethanediylbis(oxy-2,1-ethanediyl) dinitrate	97.4	CH_3CN:CH_3OH	100	Kinesis ltd.
Trimethylolethane trinitrate	TMETN	2-Methyl-3-(nitrooxy)-2-[(nitrooxy)methyl]propyl nitrate	98.5	CH_3CN:CH_3OH	100	Kinesis ltd.
Trinitrobenzene	TNB	1,3,5-Trinitrobenzene	97.5	CH_3CN:CH_3OH	1000	Kinesis ltd.
Trinitrotoluene[a]	TNT	2-Methyl-1,3,5-trinitrobenzene	≥ 99.9	CH_3CN:CH_3OH	1000	Kinesis ltd.
Urea		Urea	≥ 99	Solid	-	Sigma-Aldrich
Xylitol		D-Xylitol	≥ 99	Solid	-	Sigma-Aldrich

[a]Initial 18 target analytes used for method optimisation in Chapter 2
[b]HMDD standard was discontinued due to stability issues and so not included in all experiments.
[c]MEKP is a polymer with multiple oligomers rather than a single compound
N/A = Information not available

day and at 500 pg μL^{-1} or 5000 pg μL^{-1} across three consecutive days.

Limits of detection (LOD) were estimated for each ion based on the standard deviation of the response (σ) and the slope (S), as described by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guidelines for the validation of analytical procedures (**Error! Reference source not found.**) [108]. The standard deviation of the response in the lowest concentration detected in six replicate injections was used and slope was calculated from a plot of peak height versus concentration.

Equation 3.1: Limit of detection

$$LOD = \frac{3.3\,\sigma}{S}$$

In addition to estimating LODs, detection probability was measured for the ion with greatest signal-to-noise at 1 ng on column by analysing 30 injections. Peaks were detected using Xcalibur ICIS peak detection (with a default minimum peak width of 3 scans) and INCOS noise (a single pass algorithm used by the software to determine the noise level). This value is used by the ICIS peak detection algorithm. As some noise was also integrated, EICs were also visually examined to ensure peaks had a signal-to-noise ratio of more than 3:1 and were not just electronic spikes.

3.2.3.1 Investigating the number of isomers and theoretical elemental compositions

A molecular formula search of Chemspider (Royal Society of Chemistry, UK) database was performed on the 14/06/17 for 60 compounds to determine the number of isomers with the same molecular formula. Results were limited to single component structures only and isotopically labelled structures were disregarded. Since ammonium picrate contains two components (ammonium ion and picrate ion) it was excluded from this experiment. MEKP was also excluded due to the presence of multiple oligomers. The number of elemental compositions within 5 parts per million (ppm) of the exact mass of each of the 60 compounds was also determined using Thermo Xcalibur Qual browser software. The potential elemental compositions were restricted to the five elements included in the ions detected in Chapter 2, carbon, hydrogen, nitrogen, oxygen and chlorine, but the amount of each element was only restricted by the upper end of the mass range. This lead to consideration of the following elemental compositions: $C_{(0\text{-}52)}H_{(0\text{-}620)}N_{(0\text{-}44)}O_{(0\text{-}39)}Cl_{(0\text{-}17)}$.

3.2.3.2 Assessing selectivity

Selectivity between explosive compounds (n=60) was assessed by analysing extracted ion chromatograms for the exact mass (± 5 ppm) of the ions identified in Table 3.2 for all other target compounds in individual 10 ng μL^{-1} standards. False positives were only recorded if EIC peaks overlaid with the chromatographic peak of the most abundant ion of the explosive standard being analysed. This was to ensure common ions were produced by the main component of the standard and not due to the presence of closely related explosive compounds present as impurities. Again, MEKP was not included in the experiment due to the multiple oligomers. HMDD was also not included because the reference standard was discontinued before a 10 ng μL^{-1} standard was analysed.

A selection of matrices relevant for forensic investigations were chosen to give an initial indication of selectivity issues in matrix. The chosen matrices were natural fingermarks, indoor dirt and outdoor dirt, n=6 for each matrix. Natural fingermarks from six donors were collected on glass coverslips, extracted in 50:50 methanol:water and injected without any further sample preparation, as described in Chapter 2, Section 2.2.5. The six indoor dirt samples were collected from six different surfaces including: a laboratory floor, a window sill, a carpeted floor, an office shelf, a kitchen counter and a kitchen hob. Similarly, the six outdoor dirt samples were collected in central London from a range of surfaces including: a Perspex sign, a poster, a window sill, a metal gate, a paving slab and a wooden door. All dirt samples were collected using moistened cotton wool swabs, following an existing protocol for sample collection and extraction (unpublished work). Before sample collection, cotton swabs were placed in a 15 mL glass vial with 5 mL of 50:50 ethanol:water. Excess solvent was left in the vial and the moistened swab was used to systematically wipe the required area before being returned to the vial. The swabs were agitated within the solvent for 2 minutes before the resulting solution was drawn up through the cotton swab using a glass Pasteur pipette and placed into a new vial. The extraction process was then repeated with a second 5 mL of 50:50 ethanol:water and the two extracts were combined in the same vial. The swab extracts were then concentrated using solid phase extraction (SPE). ISOLUTE ENV + 100 mg/6 mL SPE cartridges (Biotage, Uppsala, Sweden) were initially conditioned with 1 mL of 50:50 ethanol:water. Swab extracts were then loaded onto the cartridges under gravity before elution in 1 mL of acetonitrile and analysis by LC-HRMS.

3.2.4 Passive vapour sampling of an explosives magazine

As part of a project developing “Novel passive samplers for semi-targeted explosive vapour screening” at KCL, funded by the Cross-Government Innovative Research Call in Explosives and Weapons Detection 2013 initiative; passive samplers were prepared by coating Nomex swabs (DSA Detection, MA, USA) with a 40 mg mL^{-1} solution of Tenax TA (60-80 mesh, Sigma-Aldrich) in dichloromethane (Sigma-Aldrich) and allowed to air dry as reported by McEneff et al. [109]. Passive samplers were then hung from the roof of a licenced explosive magazine and exposed for 24 hours. The contents of the explosives magazine were unknown to prevent biasing the experiment and to emulate a true screening case. Following exposure, samplers were extracted in 0.5 mL methanol and stored at -20 °C in crimp-capped LC vials until analysis, again as reported by McEneff et al. [109].

3.2.5 Data analysis

The majority of graphs were plotted in Excel. For normalised extracted ion chromatograms (EICs), the EICs were extracted in Xcalibur Qual Browser with 5 ppm mass accuracy from the exact *m/z* of the ion and no smoothing was applied. They were then normalised in Excel to the largest and smallest intensity across the full mass chromatogram. For the HCD comparison experiment, two EICs were extracted, one for each of the two MS scan types, and both were normalised to the largest intensity in either scan type, to facilitate comparison between the two scan types. SPSS was used for the elemental compositions boxplot (Figure 3.12), the one-way multivariate analysis of variance (MANOVA) and Tukey’s post-hoc test. Ion ratios were calculated in excel, using EIC peak heights with 7-point gaussian smoothing applied in Xcalibur Qual Browser.

3.3 Results and discussion

3.3.1 HRMS method development

It was not possible to optimise a method for non-target compounds. Instead an LC-HRMS method was optimised for 18 representative compounds in Chapter 2 before, the generalisable method performance of a wider set of compounds was assessed here. Achieving simultaneous ionization of a wide range of compounds is key for the development of a universal screening method, but a significant challenge for explosives analysis. Even being able to simultaneously ionize the 18 initial target compounds was a significant challenge addressed in Chapter 2. Here, the ability to ionise and detect additional explosive compounds was also considered and assessed. Additionally, fragmentation which was not addressed in Chapter 2 was considered.

When mass spectrometry is used as a confirmatory identification technique, typically a combination of fragments and molecular ions are used, either through full-scan electron impact experiments where fragments are produced during the ionisation process or tandem MS experiments looking at multiple transitions. In the case of the Exactive, the HRMS instrument used for this study, it was not possible to perform tandem mass spectrometry experiments such as product ion scanning where the first mass analyser selects a precursor ion which is then fragmented before the second mass analyser determines the *m/z* of the product ions. There are other HRMS instruments available (e.g. QExactive and QTOF) which can perform tandem MS experiments. However, in selecting precursor ions for fragmentation, detection becomes limited to a predetermined set of target compound and the potential for suspect and non-target screening is lost. Additionally, the quadrupole mass analyser used to select precursor ions is a low-resolution mass analyser and so only the fragment ions are separated with high resolution. A separate MS scan, with no fragmentation, would be required to determine the accurate mass of the selected precursor ions.

Orbitrap HRMS instruments, such as the Exactive used in this study, can perform all ion fragmentation (AIF) MS experiments, either by in-source collision induced dissociation (CID) or by using higher-energy C-trap dissociation (HCD). Alternating between full-scan and AIF MS experiments throughout an LC-MS run allows precursor and product ions that elute at the same retention time to be linked. However, this is not the same as product ion scans in tandem mass spectrometry (where specific precursor ions are selected by *m/z* for fragmentation [67]) and coeluting compounds could also be responsible for fragment ions produced by AIF.

Use of HCD fragmentation was considered in this study. Figure 3.1 shows the mass spectra recorded with and without HCD fragmentation for three explosives: a nitramine (RDX), a nitrate ester (PETN) and a nitroaromatic (TNT). Even in the full-scan mass spectra (Figure 3.1 (i)) fragment ions were detected for all three explosives. Extracted ion chromatograms are also shown for these three explosives in Figure 3.2, confirming that the ions detected in the mass spectra are associated with the chromatographic peak and not just background ions present in the mobile phase. Figure 3.2 also shows that for all the fragment ions detected for RDX and PETN, peak intensity was at least four times greater in the full-scan spectra (solid line) than when the HCD cell was used for fragmentation (dashed line). For TNT, an increase in intensity was seen for the smaller fragments when the HCD cell was used. However, all TNT fragments were also detected in the full-scan spectra and the two largest fragments (*m/z* 210.0153 and 197.0201 due to $[M\text{-}OH]^-$ and $[M\text{-}NO]^-$) were still more abundant in the full-scan spectra. Therefore, it was not necessary to use the HCD cell to induce fragmentation and instead of switching between full-scan and all ion fragmentation, which reduced the number of data point per chromatographic peak, for this study all analysis was performed in full-scan mode.

The presence of fragment ions in the full-scan spectra indicated that in-source fragmentation was occurring without introducing any additional energy. Despite both being considered soft ionisation techniques, in-source fragmentation is more common with APCI than ESI. This is because ions are more likely to have excess energy in APCI, which can lead to fragmentation, than in ESI where ions are formed in solution and so collisional cooling occurs before they reach the gas phase [110]. Unintended in-source fragmentation reduces the abundance of molecular ions, (de)protonated molecules and adduct ions. This was particularly true for nitrate esters such as PETN where the nitrate ion (*m/z* 61.9884) dominated the mass spectrum (Figure 3.1 B(i)). As discussed in Chapter 2, Section 2.3.2.1.1, autotune was used to optimise the ion optics for this study. Further optimisation of lens voltages might reduce (or increase) the amount of fragmentation and could be explored in future work, especially for targeted methods. However, the in-source fragmentation also provided additional structural information which could aid identification and no further optimisation of lens voltages was performed.

The presence of two chromatographic peaks in Figure 3.2 A and B highlighted the problem of common ions with explosives. For example, HMX produced a peak in the EICs of all the RDX ions (Figure 3.2 A) and ETN shared fragment ions with PETN (Figure 3.2 B). This demonstrated

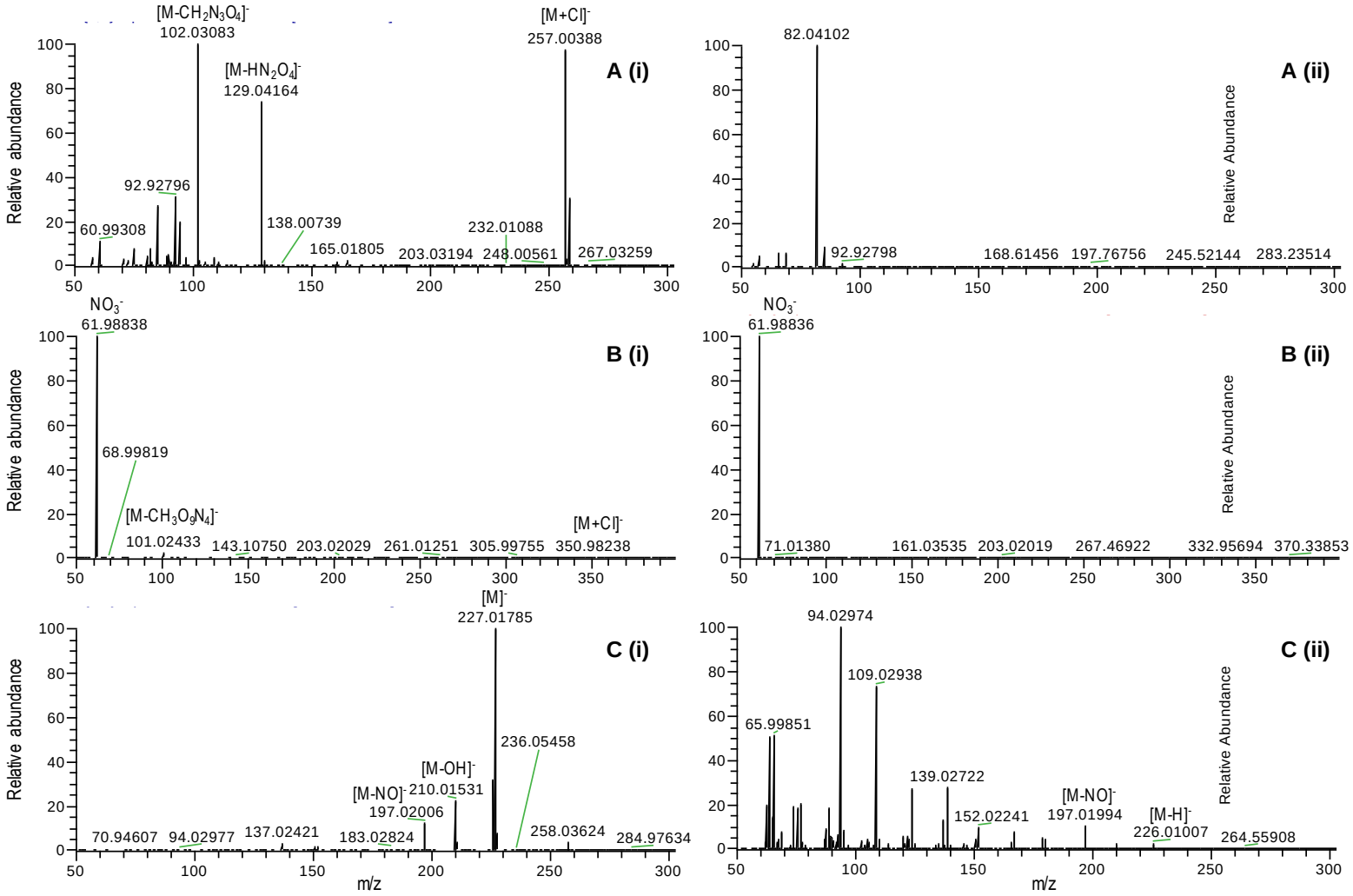


Figure 3.1: LC-HRMS mass spectra for RDX (A), PETN (B) and TNT (C), with full-scan MS spectra shown in (i) and all ion fragmentation with HCD collision energy of 20 shown in (ii).

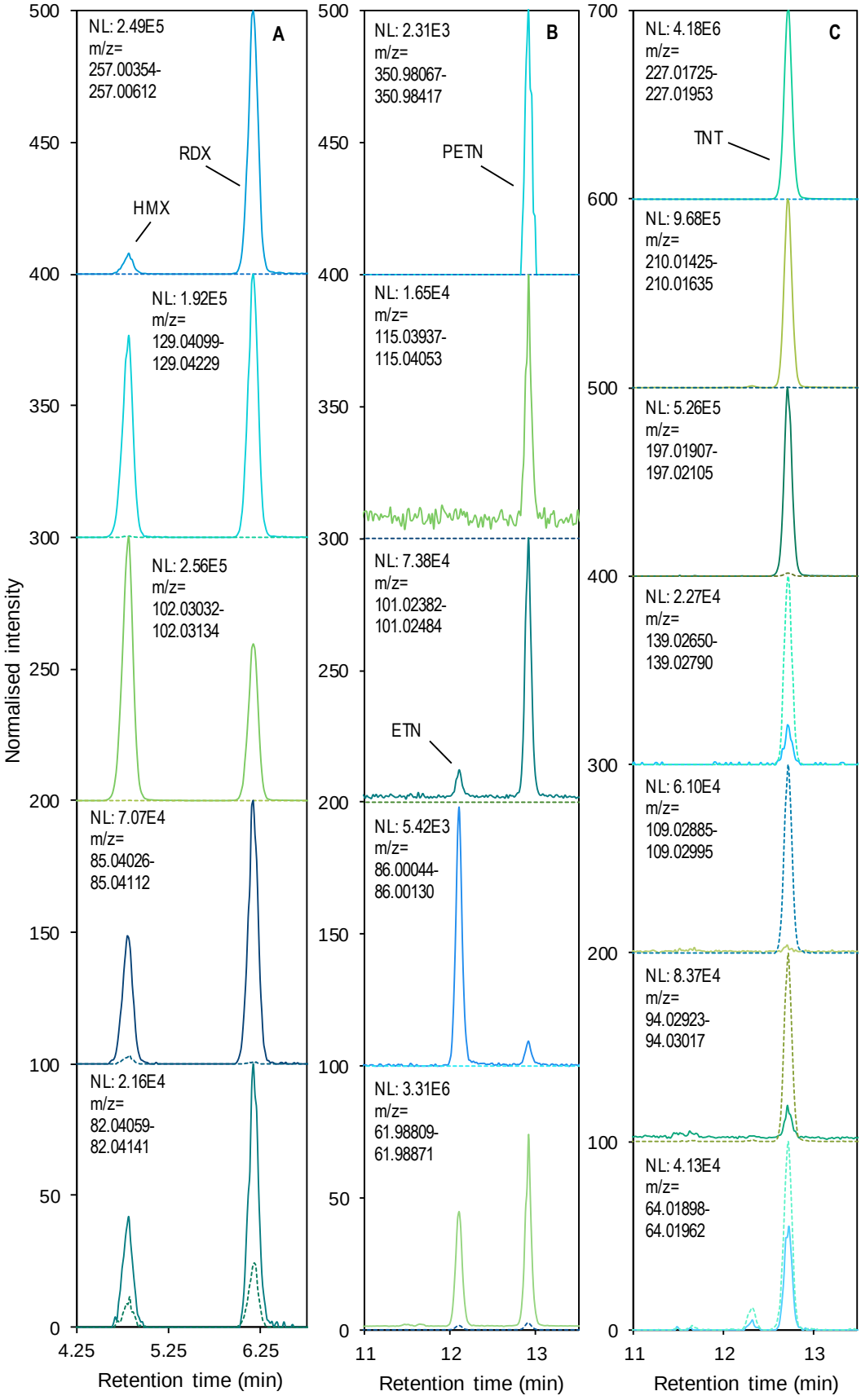


Figure 3.2: Effect of all ion fragmentation (AIF) on extracted ion chromatograms for RDX (A), PETN (B) and TNT (C) in a 10 ng μL^{-1} mixed standard solution. Solid lines represent full-scan MS and dashed lines indicate AIF with a HCD cell collision energy of 20.

the importance of developing a good chromatographic separation in Chapter 2 and the need to assess selectivity between explosive compounds here. Equally, for suspect screening applications, chromatographic selectivity is very important to help identify new compounds through retention time prediction.

3.3.2 Ions detected by LC-(APCI)-HRMS

A range of different types of ions were detected by LC-(APCI)-HRMS, as shown in Table 3.2. For example, TNT formed the molecular ion, 2,4-DNT the deprotonated molecule, HMTD the protonated molecule, RDX the chloride adduct and TATP the ammonium adduct. Various fragment ions were also produced by in-source fragmentation. For the purpose of this thesis, "Ion 1" and "Ion 2" refer to the ions with the largest and second largest signal/noise in a 10 ng μL^{-1} standard. "Ion 3" was then selected to ensure inclusion of both a fragment ion and a molecular ion/(de)protonated molecule/adduct ion which contained the intact molecule. Therefore, in the case of TNT where Ions 1 and 2 were the molecular ion and deprotonated molecule, ion 3 was selected as the most abundant fragment ion. Conversely, for nitroglycerin (NG) the chlorine adduct was selected as Ion 3 since the two most abundant ions were both fragment ions. For RDX, the two most abundant ions included both the chloride adduct and a fragment ion and so Ion 3 was simply the third most abundant ion. Signal-to-noise ratio of the chromatographic peak from the EIC was used to select Ions 1 and 2, rather than absolute intensity in the mass spectrum, to take account of background intensities which ranged from essentially no noise to in the order of 10^6 for the nitrate ion, making absolute intensities misleading.

Figure 3.3 shows the signal-to-noise ratio in a 10 ng μL^{-1} standard for the ions identified as 'Ion 1', 'Ion 2'and 'Ion 3' for 60 compounds. HMDD was discontinued for sale before it could be analysed at 10 ng μL^{-1} and MEKP, discussed in more detail in Section 3.3.2.4, was excluded from the majority of experiments due to the presence of multiple oligomers. The instrumental method sensitivity varied widely between compounds and often there was a large difference in the sensitivity between 'Ion 1' and 'Ion 3'. In fact, it was not always possible to find three ions including both a fragment ion and a molecular ion/(de)protonated molecule/adduct ion. In the case of diphenylamine (DPA) and dinitroaniline (3,5-DNA), use of the HCD cell might have helped produce fragment ions; but the HCD cell could not help with urea or nitromethane which have molecular weights of approximately 60 and 61 Da, meaning that any fragments produced would have been too small to detect. Overall, for 54 of the 61 compounds analysed, at least three ions

Table 3.2: Ions detected by LC-(APCI)-HRMS method including proposed identity and mass accuracy (n=61 compounds, MEKP oligomers not included).

Analyte	Ion 1[a]				Ion 2[b]				Ion 3[c]			
	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)
1,2-DAP	75.09156	$C_3H_{11}N_2^+$	$[M+H]^+$	-1.52	n.d.	-	-	-	n.d.	-	-	-
1,2-DNB	168.01796	$C_6H_4O_4N_2^-$	$[M]^-$	1.81	138.01993	$C_6H_4O_3N^-$	$[M-NO]^-$	1.90	154.01488	$C_6H_4O_4N^-$	$[M-N]^-$	1.93
1,2-DNG	216.98737	$C_3H_6O_7N_2Cl^-$	$[M+Cl]^-$	2.08	61.98839	NO_3^-	$[NO_3]^-$	0.36	172.00218	$C_3H_7O_5NCl^-$	$[M-NO_2+HCl]^-$	2.00
1,3-DNB	168.01727	$C_6H_4O_4N_2^-$	$[M]^-$	-2.30	138.01938	$C_6H_4O_3N^-$	$[M-NO]^-$	-2.08	167.00951	$C_6H_3O_4N_2^-$	$[M-H]^-$	-1.93
1,3-DNG	61.98836	NO_3^-	$[NO_3]^-$	-0.12	216.98721	$C_3H_6O_7N_2Cl^-$	$[M+Cl]^-$	1.34	172.00212	$C_3H_7O_5NCl^-$	$[M-NO_2+HCl]^-$	1.66
1-MNG	91.04012	$C_3H_7O_3^-$	$[M-NO_2]^-$	0.56	172.00201	$C_3H_7O_5NCl^-$	$[M+Cl]^-$	1.02	61.98831	NO_3^-	$[NO_3]^-$	-0.93
2,4-DA-6-NT	168.07683	$C_7H_{10}O_2N_3^+$	$[M+H]^+$	0.46	166.06082	$C_7H_8O_2N_3^+$	$[M-H_2]^+$	-1.71	121.0762	$C_6H_8O_2N_3^+$	$[M-HNO_2]^+$	1.20
2,4-DNT	181.02571	$C_7H_5O_4N_2^-$	$[M-H]^-$	1.26	182.03329	$C_7H_6O_4N_2^-$	$[M]^-$	-0.09	165.03072	$C_7H_5O_3N_2^-$	$[M-OH]^-$	0.93
2,6-DA-4-NT	168.07684	$C_7H_{10}O_2N_3^+$	$[M+H]^+$	0.52	180.07698	$C_8H_{10}O_2N_3^+$	$[M+CH]^+$	1.26	122.08385	$C_6H_9O_2N_3^+$	$[M-NO_2]^+$	0.00
2,6-DNT	182.03363	$C_7H_6O_4N_2^-$	$[M]^-$	1.78	152.03558	$C_7H_6O_3N^-$	$[M-NO]^-$	1.73	181.02589	$C_7H_5O_4N_2^-$	$[M-H]^-$	2.26
2-A-4,6-DNT	196.03633	$C_7H_6O_4N_3^-$	$[M-H]^-$	-0.26	232.01302	$C_7H_7O_4N_3^-$	$[M+Cl]^-$	-0.20	180.04146	$C_7H_6O_3N_3^-$	$[M-OH]^-$	-0.03
2-MNG	172.00166	$C_3H_7O_5NCl^-$	$[M+Cl]^-$	-1.02	91.04008	$C_3H_7O_3^-$	$[M-NO_2]^-$	0.13	61.98844	NO_3^-	$[NO_3]^-$	1.17
2-NT	136.04053	$C_7H_6O_2N^-$	$[M-H]^-$	0.94	121.05344	$C_7H_7ON^-$	$[M-O]^-$	1.05	120.0456	$C_7H_5ON^-$	$[M-OH]^-$	0.94
3,4-DNT	182.03334	$C_7H_6O_4N_2^-$	$[M]^-$	0.18	152.0354	$C_7H_6O_3N^-$	$[M-NO]^-$	0.54	181.02563	$C_7H_5O_4N_2^-$	$[M-H]^-$	0.82
3,5-DNA	183.02884	$C_6H_5O_4N_3^-$	$[M]^-$	1.55	182.02107	$C_6H_4O_4N_3^-$	$[M-H]^-$	1.86	n.d.	-	-	-
3-NT	121.05334	$C_7H_7ON^-$	$[M-O]^-$	0.23	137.04819	$C_7H_7O_2N^-$	$[M]^-$	-0.27	138.05602	$C_7H_8O_2N^-$	$[M+H]^-$	-0.24
4-A-2,6-DNT	196.03621	$C_7H_6O_4N_3^-$	$[M-H]^-$	-0.87	232.01344	$C_7H_7O_4N_3^-$	$[M+Cl]^-$	1.61	167.04602	$C_7H_7O_3N_2^-$	$[M-NO]^-$	-1.18
4-NT	136.0405	$C_7H_6O_2N^-$	$[M-H]^-$	0.72	120.04552	$C_7H_5ON^-$	$[M-OH]^-$	0.27	121.05342	$C_7H_7ON^-$	$[M-O]^-$	0.89
Aspirin	198.07619	$C_9H_{12}O_4N^+$	$[M+NH_4]^+$	0.53	n.d.	-	-	-	n.d.	-	-	-
DADP	89.05962	$C_4H_9O_2^+$	$[M-C_2H_3O_2]^+$	-0.97	91.03892	$C_3H_7O_3^+$	$[M-C_3H_5O]^+$	-0.56	n.d.	-	-	-
DDNP	183.00497	$C_6H_3O_5N_2^-$	$[M+H-N_2]^-$	1.17	213.0154	$C_7H_5O_6N_2^-$	$[M+OCH_3-N_2]^-$	0.36	244.97188	$C_6H_2O_5N_4Cl^-$	$[M+Cl]^-$	-0.21
DNSA	226.99460	$C_7H_3O_7N_2^-$	$[M-H]^-$	0.04	210.99968	$C_7H_3O_6N_2^-$	$[M-OH]^-$	0.03	228.00224	$C_7H_4O_7N_2^-$	$[M]^-$	-0.77

Table 3.2 (Continued): Ions detected by LC-(APCI)-HRMS method including proposed identity and mass accuracy (n=61 compounds, MEKP oligomers not included).

Analyte	Ion 1[a]				Ion 2[b]				Ion 3[c]			
	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)
DEDPU	269.1651	$C_{17}H_{21}ON_2^+$	$[M+H]^+$	0.97	120.04443	$C_7H_6ON^+$	$[M-C_{10}H_{14}N]^+$	0.33	148.07552	$C_9H_{10}ON^+$	$[M-C_8H_{10}N]^+$	-1.15
DEGDN	61.98846	NO_3^-	$[NO_3]^-$	1.49	103.04008	$C_4H_7O_3^-$	$[M-HO_4N_2]^-$	0.11	231.00261	$C_4H_8O_7N_2Cl^-$	$[M+Cl]^-$	0.18
DMDPU	241.13382	$C_{15}H_{17}ON_2^+$	$[M+H]^+$	1.16	134.06027	$C_8H_8ON^+$	$[M-C_7H_8N]^+$	1.71	106.06505	$C_7H_8N^+$	$[M-C_8H_8NO]^+$	-0.71
DMNB	194.11378	$C_6H_{16}O_4N_3^+$	$[M+NH_4]^+$	1.27	74.05998	$C_3H_8ON^+$	$[M-C_3H_4O_3N]^+$	-0.82	99.08051	$C_6H_{11}O^+$	$[M-HO_3N_2]^+$	0.69
DPA	170.0965	$C_{12}H_{12}N^+$	$[M+H]^+$	0.41	n.d.	-	-	-	n.d.	-	-	-
EGDN	61.98828	NO_3^-	$[NO_3]^-$	-1.41	n.d.	-	-	-	n.d.	-	-	-
Erythritol	157.02744	$C_4H_{10}O_4Cl^-$	$[M+Cl]^-$	0.77	121.05072	$C_4H_9O_4^-$	$[M-H]^-$	0.71	102.03231	$C_4H_6O_3^-$	$[M-H_4O]^-$	0.65
ETN	115.00348	$C_4H_3O_4^-$	$[M-H_3N_4O_8]^-$	-1.77	86.00088	$C_3H_2O_3^-$	$[M-CH_4O_9N_4]^-$	-0.74	336.96756	$C_4H_6O_{12}N_4Cl^-$	[M+Cl]-	-0.27
HMDD	177.08704	$C_6H_{13}O_4N_2^+$	$[M+H]^+$	0.31	129.0659	$C_5H_9O_2N_2^+$	$[M-CH_3O_2]^+$	0.35	113.07096	$C_5H_9ON_2^+$	$[M-CH_3O_3]^+$	0.18
HMTA	141.1136	$C_6H_{13}N_4^+$	$[M+H]^+$	0.90	106.08631	$C_4H_{12}O_2N^+$	$[M-C_2N_2+O_2]^+$	0.51	150.11264	$C_6H_{16}O_3N^+$	$[M-N_3+H_4O_3]^+$	1.13
HMTD	145.06075	$C_5H_9O_3N_2^+$	$[M-CH_3O_3]^+$	-0.13	179.06628	$C_5H_{11}O_5N_2^+$	$[M-CHO]^+$	0.17	209.07692	$C_6H_{13}O_6N_2^+$	$[M+H]^+$	0.51
HMX	331.01703	$C_4H_8O_8N_8Cl^-$	$[M+Cl]^-$	3.33	102.03102	$C_2H_4O_2N_3^-$	$[M-C_2H_4N_5O_6]^-$	1.17	284.01559	$C_4H_7O_6N_7Cl^-$	$[M-HNO_2+Cl]^-$	1.39
HNDPA	438.99979	$C_{12}H_5O_{12}N_7^-$	$[M]^-$	-0.87	437.99216	$C_{12}H_4O_{12}N_7^-$	$[M-H]^-$	-0.43	376.00452	$C_{12}H_4O_9N_6^-$	$[M-HNO_3]^-$	-0.02
Mannitol	217.04832	$C_6H_{14}O_6Cl^-$	$[M+Cl]^-$	-0.61	181.07172	$C_6H_{13}O_6^-$	$[M-H]^-$	-0.24	163.0611	$C_6H_{11}O_5^-$	$[M-H_3O]^-$	-0.60
MHN	251.0153	$C_6H_7O_9N_2^-$	$[M-HN_4O_9]^-$	-1.62	314.01089	$C_6H_8O_{12}N_3^-$	$[M-N_3O_6]^-$	-1.47	513.97681	$C_6H_8O_{21}N_7^-$	$[M+NO_3]^-$	-2.08
NB	107.03773	$C_6H_5ON^-$	$[M-O]^-$	0.63	123.03263	$C_6H_5O_2N^-$	$[M]^-$	0.43	138.01968	$C_6H_4O_3N^-$	$[M-H+O]^-$	0.09
NG	86.00093	$C_3H_2O_3^-$	$[M-H_3N_3O_6]^-$	-0.16	61.98841	NO_3^-	$[NO_3]^-$	0.68	261.97183	$C_3H_5O_9N_3Cl^-$	$[M+Cl]^-$	-0.65
NM	60.0092	$CH_2O_2N^-$	$[M-H]^-$	1.62	n.d.	-	-	-	n.d.	-	-	-
NQ	105.04071	$CH_5O_2N_4^+$	$[M+H]^+$	0.07	122.06736	$CH_8O_2N_5^+$	$[M+NH_4]^+$	0.89	75.04263	$CH_5ON_3^+$	$[M-NO+H]^+$	-1.11
PA	227.98943	$C_6H_2O_7N_3^-$	$[M-H]^-$	-1.74	211.99448	$C_6H_2O_6N_3^-$	$[M-OH]^-$	-2.03	228.9979	$C_6H_3O_7N_3^-$	$[M]^-$	1.09
PETN	61.98839	NO_3^-	$[NO_3]^-$	0.36	350.98288	$C_5H_8O_{12}N_4Cl^-$	$[M+Cl]^-$	-1.20	101.02435	$C_4H_5O_3^-$	$[M-CH_3O_9N_4]^-$	-0.68
PGDN	61.98839	NO_3^-	$[NO_3]^-$	0.36	118.01488	$C_3H_4O_4N^-$	$[M-H_2NO_2]^-$	2.52	200.99277	$C_3H_6O_6N_2Cl^-$	$[M+Cl]^-$	3.82
Picramic acid	198.01532	$C_6H_4O_5N_3^-$	$[M-H]^-$	-1.64	182.02109	$C_6H_4O_4N_3^-$	$[M-OH]^-$	1.93	199.02298	$C_6H_5O_5N_3^-$	$[M]^-$	-2.47

Table 3.2 (Continued): Ions detected by LC-(APCI)-HRMS method including proposed identity and mass accuracy (n=61 compounds, MEKP oligomers not included).

Analyte	Ion 1[a]				Ion 2[b]				Ion 3[c]			
	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)	Accurate mass	Elemental composition	Proposed ion	Δ (ppm)
Picrate	227.98967	$C_6H_2O_7N_3^-$	$[M]^-$	-0.68	211.99474	$C_6H_2O_6N_3^-$	$[M-O]^-$	-0.80	228.99736	$C_6H_3O_7N_3^-$	$[M+H]^-$	-1.27
PYX	621.00659	$C_{17}H_7O_{16}N_{11}^-$	$[M]^-$	-1.91	619.99921	$C_{17}H_6O_{16}N_{11}^-$	$[M-H]^-$	-1.20	210.0027	$C_6H_2O_5N_4^-$	$[M-C_{11}H_5O_{11}N_7]^-$	-1.76
RDX	257.00494	$C_3H_6O_6N_6Cl^-$	$[M+Cl]^-$	2.50	102.0311	$C_2H_4O_2N_3^-$	$[M-CH_2N_3O_4]^-$	1.95	129.04213	$C_3H_5O_2N_4^-$	$[M-HN_2O_4]^-$	2.56
R-Salt	209.01941	$C_3H_6O_3N_6Cl^-$	$[M+Cl]^-$	-0.65	144.05257	$C_3H_6O_2N_5^-$	$[M-NO]^-$	-0.89	86.0359	$C_2H_4ON_3^-$	$[M-CH_2O_2N_3]^-$	-0.99
Salicylic acid	137.02446	$C_7H_5O_3^-$	$[M-H]^-$	0.30	93.03456	$C_7H_6O_3Cl^-$	$[M+Cl]^-$	-0.31	173.0011	$C_6H_5O^-$	$[M-CHO_2]^-$	-0.01
SHN	251.0154	$C_6H_7O_9N_2^-$	$[M-HN_4O_9]^-$	-1.22	314.01102	$C_6H_8O_{12}N_3^-$	$[M-N_3O_6]^-$	-1.05	513.97711	$C_6H_8O_{21}N_7^-$	$[M+NO_3]^-$	-1.50
Sorbitol	217.04837	$C_6H_{14}O_6Cl^-$	$[M+Cl]^-$	-0.38	181.07176	$C_6H_{13}O_6^-$	$[M-H]^-$	-0.02	163.0611	$C_6H_{11}O_5^-$	$[M-H_3O]^-$	-0.60
TATB	257.02789	$C_6H_5O_6N_6^-$	$[M-H]^-$	1.10	240.02499	$C_6H_4O_5N_6^-$	$[M-H_2O]^-$	0.51	241.03287	$C_6H_5O_5N_6^-$	$[M-HO]^-$	0.74
TATP	89.05963	$C_4H_9O_2^+$	$[M-C_5H_9O_4]^+$	-0.86	91.03892	$C_3H_7O_3^+$	$[M-C_6H_{11}O_3]^+$	-0.56	240.14447	$C_9H_{22}O_6N^+$	$[M+NH_4]^+$	1.27
TEGDN	258.09302	$C_6H_{16}O_8N_3^+$	$[M+NH_4]^+$	-0.67	151.09634	$C_6H_{15}O_4^+$	$[M-N_2O_4+H_3]^+$	-0.97	196.08145	$C_6H_{14}O_6N^+$	$[M-NO_2+2H]^+$	-0.59
Tetryl	241.0218	$C_7H_5O_6N_4^-$	$[M-NO_2]^-$	1.41	212.01907	$C_6H_4O_5N_4^-$	$[M-CHNO_3]^-$	1.65	321.98364	$C_7H_5O_8N_5Cl^-$	$[M+Cl]^-$	1.27
TMETN	85.0296	$C_4H_5O_2^-$	$[M-CH_4O_7N_3]^-$	1.13	290.0033	$C_5H_9O_9N_3Cl^-$	$[M+Cl]^-$	0.00	245.01801	$C_5H_{10}O_7N_2Cl^-$	$[M-NO_2+HCl]^-$	-0.85
TNB	213.00252	$C_6H_3O_6N_3^-$	$[M]^-$	-1.01	183.00456	$C_6H_3O_5N_2^-$	$[M-NO]^-$	-1.02	244.02081	$C_7H_6O_7N_3^-$	$[M+CH_3O]^-$	-1.29
TNT	227.01862	$C_7H_5O_6N_3^-$	$[M]^-$	1.03	226.01089	$C_7H_4O_6N_3^-$	$[M-H]^-$	1.46	197.02061	$C_7H_5O_5N_2^-$	$[M-NO]^-$	1.08
Urea	61.03968	$CH_5O_2^+$	$[M+H]^+$	0.67	78.06612	$CH_8O_2N^+$	$[M+NH_4]^+$	-0.88	n.d.			
Xylitol	187.03787	$C_5H_{12}O_5Cl^-$	$[M+Cl]^-$	-0.09	151.06119	$C_5H_{11}O_5^-$	$[M-H]^-$	-0.06	132.04276	$C_5H_8O_4^-$	$[M-H_3O]^-$	-0.37

[a].Ion giving EIC peak with largest S:N

[b].Ion giving EIC peak with second largest S:N

[c]molecular ion/(de)protonated molecule/adduct ion with largest S:N if 1 and 2 are both fragments, fragment ion with largest S:N if 1 and 2 are both molecular ion/(de)protonated molecule/adduct ions or ion with third largest S:N if ions 1 and 2 are a molecular ion/(de)protonated molecule/adduct and a fragment ion

n.d. not detected

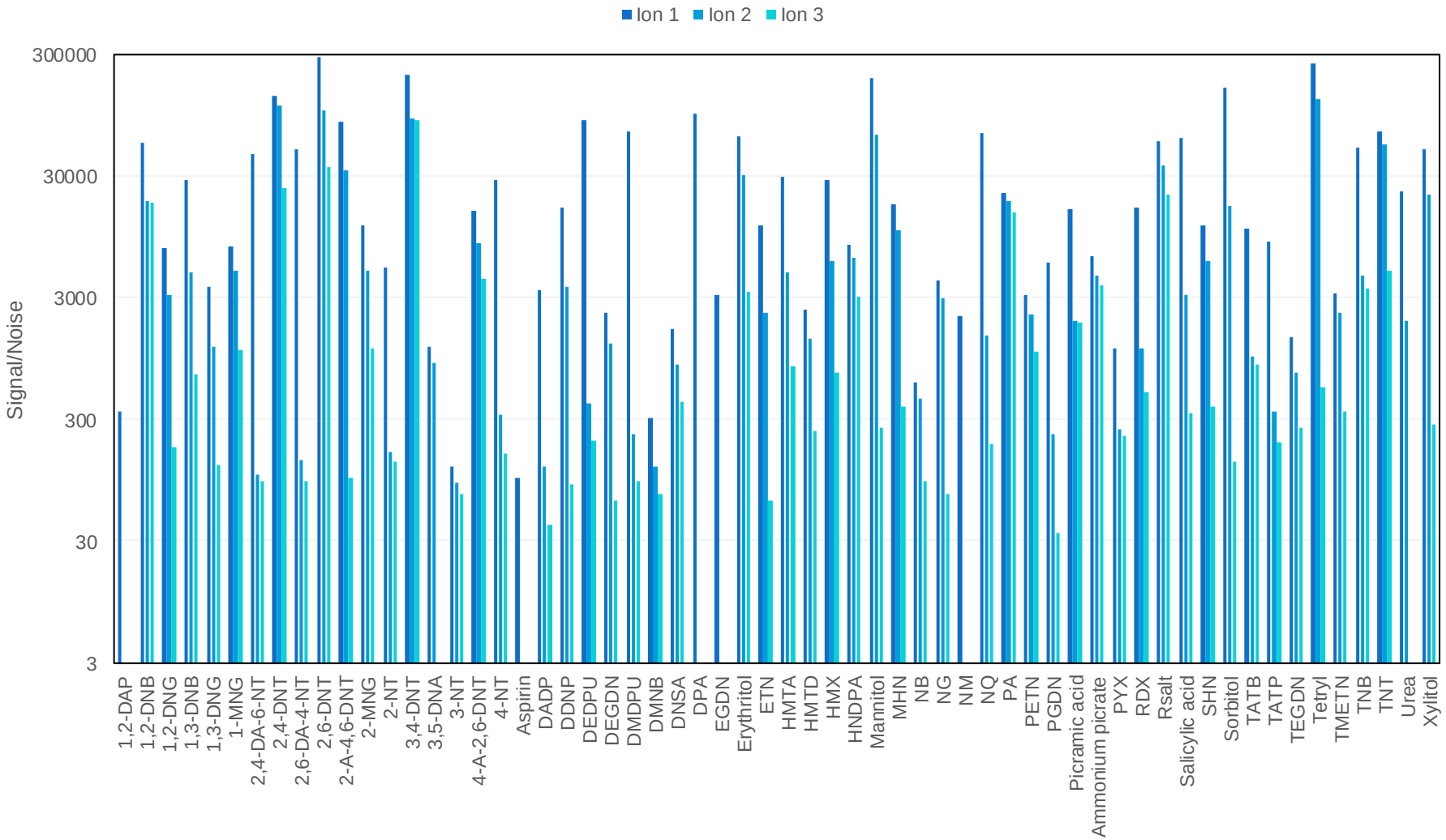


Figure 3.3: Signal-to-noise ratio of the ions identified as Ion 1, 2 and 3 for each compound in a 10 ng μL^{-1} standard (n=60, HMDD and MEKP oligomers not included).

(including the molecular ion/(de)protonated molecule/adduct ion and a fragment ion) could be detected at the highest concentration (10 ng μL^{-1}) and at least one ion was detected for all 61 compounds.

In addition to the recorded accurate *m/z*, Table 3.2 includes the proposed ion, elemental composition and mass difference (Δ) in parts per million (ppm) between the measured accurate mass and calculated exact mass. Where fragmentation pathways have been previously studied and fragment ion structures have already been identified, fragment ions were easily identified. However, fragmentation pathways were not always available in the literature and in some instances multiple elemental compositions had an exact mass within 5 ppm. Further experiments, such as using isotopically labelled starting materials, would be required to identify fragment ions in those cases.

In general, the nitro-explosives were detected in negative mode APCI while the peroxides and amines were detected in positive mode APCI. This meant that either separate injections for each ionization mode or polarity switching were required. While polarity switching is worth considering if speed of analysis is a priority, here separate injections were used for positive and negative analysis to maximize both the number of data points per chromatographic peak and instrumental method sensitivity which was adversely affected by polarity switching (data not shown).

3.3.2.1 Nitramines

The nitramines (RDX, HMX and Tetryl) were detected as chloride adducts rather than molecular ions or deprotonated molecules. RDX and HMX shared common fragment ions, as shown in Figure 3.2. For RDX, the fragment ions detected here with an accurate *m/z* of 102.0311 and 129.04213 had a nominal mass consistent with fragments previously identified in the literature by gas chromatography coupled to mass spectrometry (GC/MS) using negative ion chemical ionisation (NICI) [111]. The measured accurate masses were also consistent (Δ = 1.95 and 2.56 ppm, Table 3.2) with the elemental compositions of the fragment ion structures reported by Florián et al. [111]. Interestingly neither of these fragment ions were reported by Xu et al. [33] or DeTata et al. [32] when they used APCI, the same ionization source used here, in their HRMS screening methods. R-Salt, a nitrosamine sometimes referred to as trinitroso-RDX or TNX, also formed a chloride adduct, but in this case $[M\text{-}NO]^-$ was the major fragment. Again, the two R-Salt fragment ions detected here had accurate masses consistent with the nominal masses and fragmentation pathways reported by Florián et al. [111]. For Tetryl, which contains both nitramine and

nitroaromatic functionality, the $[M\text{-}NO_2]^-$ ion was the most abundant ion as reported by Xu et al. [33] and DeTata et al. [32].

3.3.2.2 Nitroaromatics

The nitroaromatics are typically detected by negative mode ionization, forming both molecular ions and/or deprotonated molecules. The only exception to this were the diamino-nitrotoluenes (2,4-DA-6-NT and 2,6-DA-4-NT) which ionised in positive mode to form protonated molecules. Other than Tetryl, the only nitroaromatics detected as chloride adducts were the amino-dinitrotoluenes (2-A-4,6-DNT and 4-A-2,6-DNT) and diazodinitrophenol (DDNP). The hydroxylate isoform, $[M+H]^-$, and protonated molecule, $[M+H]^+$, were also detected for DDNP as reported by Havliková et al. [112], but here they were detected with accurate masses of 211.0109 and 211.0098 with the difference in the two masses (0.0011 Da) explained by the difference in electrons (± 0.00055 Da). The base peak for DDNP (*m/z* of 183.00479) also had the same nominal mass as the most abundant fragment ion (*m/z* 182.9, due to the loss of N_2 from the hydroxylate isoform) reported by Havliková et al. [112]. The most challenging nitroaromatics to detect were the nitrotoluenes, particularly 3-NT, and nitrobenzene (NB). Interestingly for both 3-NT and NB, an ion corresponding to the loss of oxygen, $[M\text{-}O]^-$, had the largest signal to noise ratio.

For the nitroaromatics detected in negative mode, the most common fragment ions were $[M\text{-}NO]^-$ and/or $[M\text{-}OH]^-$. Fragmentation pathways for TNT to the $[M\text{-}OH]^-$ and $[M\text{-}NO]^-$ ions have been proposed elsewhere [113] and these fragments were also detected by DeTata et al. following precursor selection and fragmentation [32]. Density functional theory calculations by Nguyen Van et al. suggested that formation of an aci-isomer of TNT radical anion, where the ortho nitro-group gained a hydrogen from the methyl group, followed by loss of a hydroxyl radical was the favoured method for formation of the $[M\text{-}OH]^-$ ion (Figure 3.4); whereas the methyl group was not involved

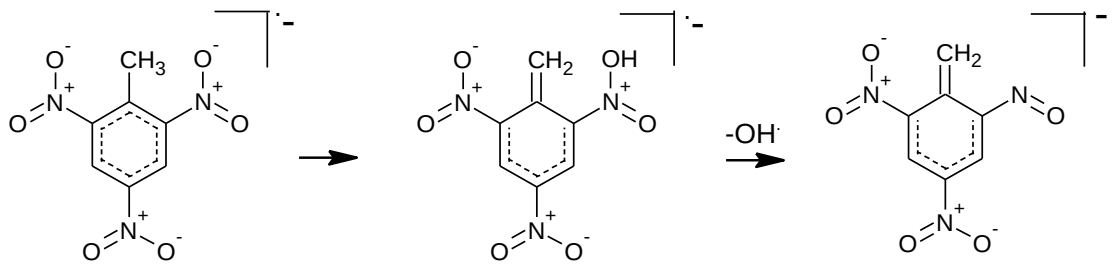


Figure 3.4: Loss of hydroxyl radical from ortho-aci form of TNT radical anion. Adapted from [113].

in formation of the $[M\text{-}NO]^-$ ion [113]. This could also explain why Table 3.2 shows that formation of the $[M\text{-}NO]^-$ fragment was favoured over the $[M\text{-}OH]^-$ fragment for 3,4-DNT and the DNB isomers, since 3,4-DNT does not contain an ortho nitro-group and the DNB isomers do not contain a methyl group. Interestingly 2,4-DNT and 2-A-4,6-DNT, which contain both an ortho and a para nitro-group, favoured formation of the $[M\text{-}OH]^-$ fragment; whereas 2,6-DNT and 4-A-2,6-DNT, which contain two ortho nitro groups, favoured formation of the $[M\text{-}NO]^-$ fragment. This suggested that the presence of a para nitro group may also aid formation of the $[M\text{-}OH]^-$ fragment. Additionally, ion ratios may be useful for distinguishing between isomers which produced fragment ions with the same *m/z* and so should be investigated.

3.3.2.3 Nitrate esters

The majority of nitrate esters (10 out of 14) were detected as chloride adducts (Table 3.2), as reported elsewhere [33,34]. The only exceptions were EGDN which was only detected as a nitrate ion, mannitol and sorbitol hexanitrate (MHN and SHN) which formed a nitrate adduct in preference and TEGDN which was detected best in positive mode and formed an ammonium adduct. DEGDN was also detected in positive mode as an ammonium adduct (*m/z*=214.06731) and fragment ions (*m/z*=122.08140 and *m/z*=169.08217), but as overall signal-to-noise was better in negative mode, the negative mode ions were used here in preference. Fragment ions had the greatest signal-to-noise ratio for all nitrate esters, except for TEGDN and the DNG isomers, despite often having greater background noise than the adduct ions. The nitrate ion (NO_3^-) was detected for all 14 nitrate esters analyzed here and has previously been used by DeTata et al. for detection of 1- and 2-MNG, 1,2- and 1,3-DNG, EGDN, PGDN, DEGDN and TEGDN [32]. Screening for the nitrate ion may therefore be useful for biased non-target screening, when searching for unknown nitrate ester peaks of interest which may otherwise be difficult to find amongst all the full-scan data collected.

The mannitol hexanitrate (MHN) sample was of unknown purity and multiple chromatographic peaks corresponding to the *m/z* for the chloride adducts and the loss of NO_2 from mannitol pentanitrate (MPN) and tetranitrate (MTN) were also detected (Figure 3.5). The sample of sorbitol hexanitrate (SHN), a stereoisomer of MHN, was also of unknown purity and again, peaks corresponding to partially nitrated sugars, in this case sorbitol dinitrate (SDN) and trinitrate (STriN) as well as the tetranitrate (STN) and pentanitrate (SPN), were also detected (Figure 3.5). The presence of the dinitrates and trinitrates of sorbitol may be due to a less efficient synthesis of

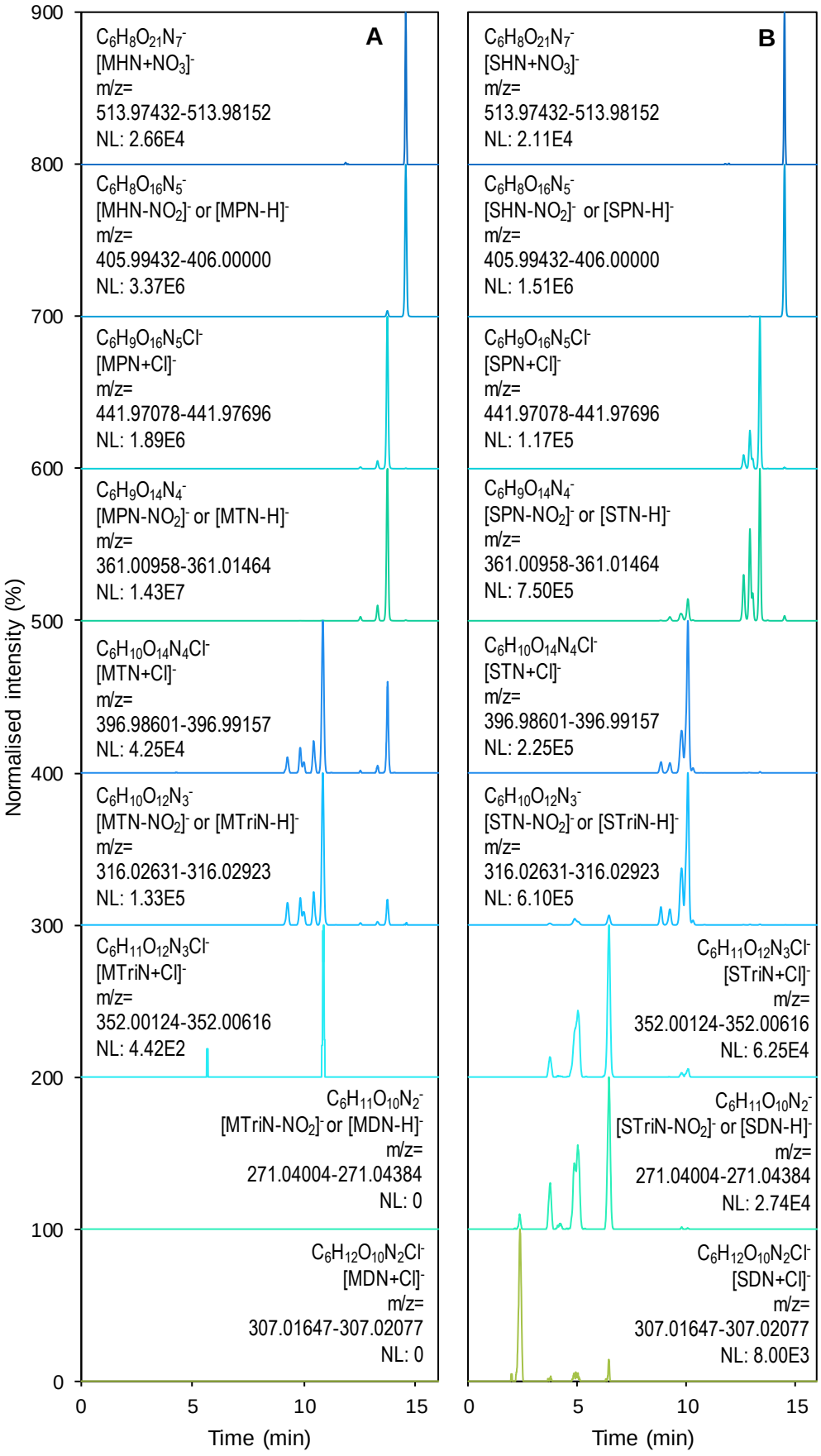


Figure 3.5: Chromatographic separation of: a) mannitol hexanitrate and b) sorbitol hexanitrate from the less nitrated sugars, pentanitrate (MPH and SPN), tetranitrate (MTN and STN) trinitrate (MTriN ansSTriN) and dinitrate (MDN and SDN).

sorbitol hexanitrate, or more degradation post synthesis, compared to the mannitol hexanitrate sample. However, as neither sample was analysed immediately after synthesis it was not possible to determine which was the main factor. Ostrinskaya et al. also reported detection of partially nitrated mannitol and sorbitol [114]. However, only one chromatographic peak was shown for MPN and SDN, STriN, STN and SPN were reported as coeluting in one peak. Furthermore, for the majority of experiments, Ostrinskaya et al. did not use any chromatography and so were unable to distinguish between in-source fragmentation of the hexanitrates and partially nitrated sugars.

Here, three peaks were detected in the EICs for MPN which has three isomers due to different locations of the hydroxyl group (Figure 3.6). Three peaks were also seen for SPN, although as SHN does not have the rotational symmetry present in MHN, theoretically there are six isomers of SPN. However, given that a chiral stationary phase was not used, separation of pairs of sterioisomers was not expected. For the same reason, it was unsurprising that mannitol and sorbitol hexanitrate were not chromatographically separated, as they are also stereoisomers. Figure 3.5 does show chromatographic separation between structural isomers. However, as individual reference materials were unavailable it was not possible to determine which peak was due to which isomer.

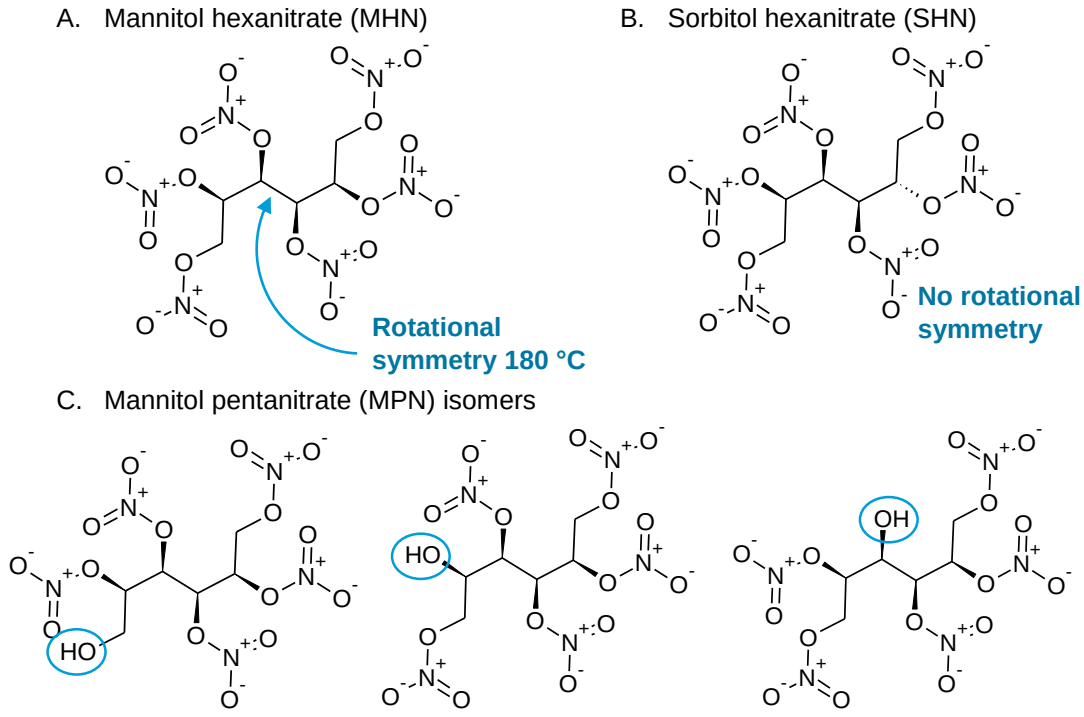


Figure 3.6: Chemical structures of A) mannitol hexanitrate, B) sorbitol hexanitrate and C) the three isomers of mannitol pentanitrate.

Chromatographic separation was also seen between MHN or SHN and most peaks corresponding to the exact *m/z* for, pentanitrate, tetranitrate, trinitrate and dinitrate ions, indicating that these were due to partially nitrated sugars rather than in-source fragmentation of MHN or SHN. However, it was not possible to conclusively identify peaks due to the partially nitrated sugars without reference standards. In some cases, peaks detected in the EICs, shown in Figure 3.5, could be due to the loss of a nitro group, a deprotonated molecule of a less nitrated sugar, or the loss of multiple nitro groups from a more-nitrated sugar. For example, in addition to $[SPN-NO_2]^-$ ions, peaks in the EIC for *m/z*: 361.01211 ± 5 ppm could also be due to $[STN-H]^-$ or $[SHN-2NO_2+H]^-$ ions which would both also have an elemental composition of $C_6H_9N_4O_{14}^-$. It was expected that the compounds with greater nitration were retained on the column for longer since each nitrate group ($-ONO_2$) replaces a more hydrophilic hydroxyl group (-OH), which unlike the nitrate group can act as both a hydrogen donor and hydrogen acceptor for hydrogen bonding with the mobile phase. While the nitrogen atom in a nitrate group has a positive charge due to the formation of four bonds, this is balanced by the negative charge on one of the oxygen atoms which is only able to form one bond, resulting in a net neutral charge. Therefore, nitrate groups are only able to form permanent dipole interactions or act as a hydrogen acceptor in hydrogen bonding. As pure reference standards were not available for the nitrated sugars, it was not possible to confirm the theory that the more nitrated sugars would be retained longer. However, in Chapter 4, predicted retention times will be determined and identification of the partially nitrated sugars will be revisited.

3.3.2.4 Peroxides

Hexamethylene triperoxide diamine (HMTD) and hexamethylene diperoxide diamine (HMDD) were the only peroxides detected as the protonated molecule. The two ions with the greatest signal-to-noise ratio for HMTD were fragment ions with similar accurate masses (145.0606 vs 145.06075 and 179.0660 vs 179.06628) to ions detected by Colizza et al. and tentatively identified [115]. Colizza et al. found that the protonated molecule of HMTD was the most abundant ion when they used an acetonitrile mobile phase but with a methanol mobile phase HMTD reacted in the gas phase to produce an ion with a *m/z* of 207.0979 ($[HMTD + H^+ + MeOH - H_2O_2]^+$) [115]; this ion was also detected in this study, although the fragment ions were more abundant. Triacetone triperoxide (TATP) and methyl ethyl ketone peroxide (MEKP) were detected as

ammonium adducts rather than protonated molecules and diacetone diperoxide (DADP) was only detected as fragment ions.

The most abundant ion for both TATP and DADP was $C_4H_9O_2^+$, with an accurate mass of 89.0596. Widmer et al. reported an ion with *m/z* 89 as the base peak for TATP in their LC-MS method over 15 years ago, but at that point the structure was unknown [96]. Recently, through LC-HRMS analysis of TATP and d_{18}-TATP, Colizza et al. showed that the TATP ion with an exact *m/z* of 89.0597 was a gas-phase reaction product of TATP with one or two methanol molecules [116].

The reaction mechanism proposed by Colizza et al. is shown in Figure 3.7. Due to the similarity in structure between DADP and TATP, it is likely that formation of the *m/z* 89.0596 ion follows a similar mechanism for DADP with the incorporation of one (Figure 3.8) or two (Figure 3.7 B) methanol molecules. The second most abundant ion seen here for TATP and DADP was $C_3H_7O_3^+$ with an accurate mass of 91.0389. Colizza et al. reported two structures for this ion (exact *m/z* 91.0390), one with and one without the incorporation of a methanol molecule [116]. Whilst Colizza et al. were also using HRMS, it is worth noting the greater mass accuracy detected here (absolute Δ < 1 ppm, for the *m/z* 89.0596 and *m/z* 91.0390 ions) compared to the mass accuracy reported by Colizza et al. for these ions (absolute Δ = 9 - 11 ppm).

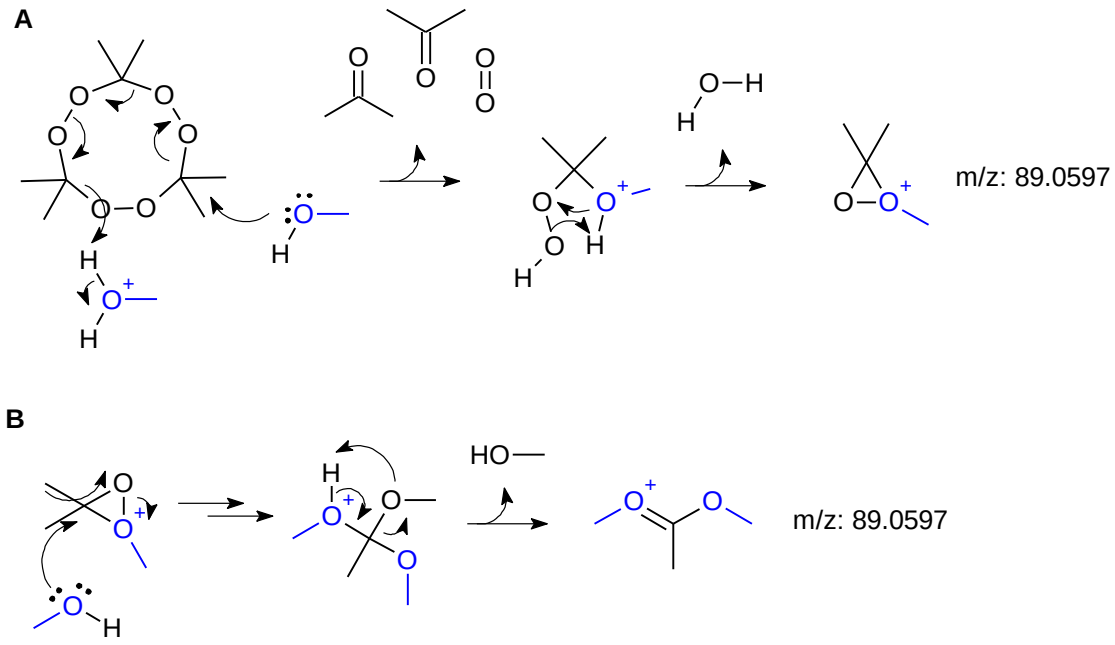


Figure 3.7: Mechanisms proposed by Colizza et al. for A) the addition of one molecule of methanol to TATP and B) the addition of a second methanol molecule [116].

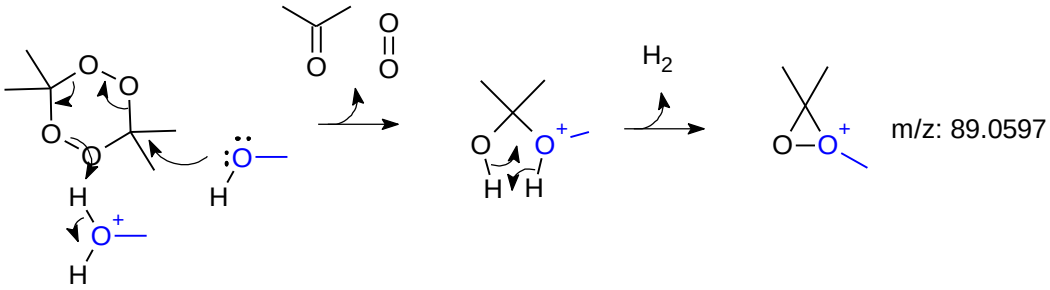


Figure 3.8: Proposed mechanism for the addition of one methanol molecule to DADP.

Methyl ethyl ketone peroxide (MEKP) was not included in Table 3.2, or the method performance experiments because it is a polymer which can form cyclic or linear oligomers of various sizes (Figure 3.9). The MEKP sample analysed here contained multiple chromatographic peaks with ions corresponding to the ammonium adduct of various oligomers of MEKP (Figure 3.10). The

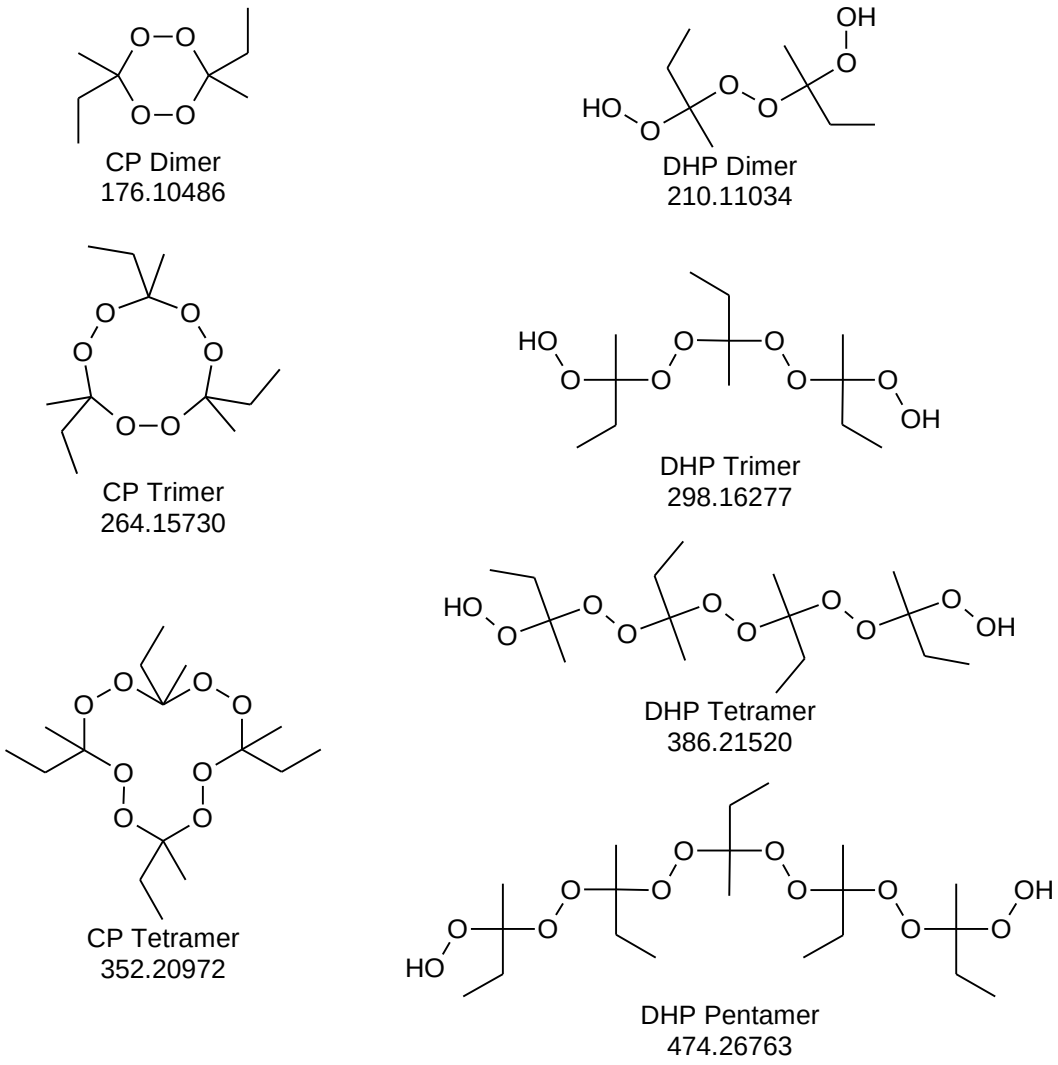


Figure 3.9: Chemical structures and monoisotopic masses for cyclic peroxide (CP) and dihydroperoxy peroxide (DHP) oligomers of MEKP, which were detected as ammonium adducts.

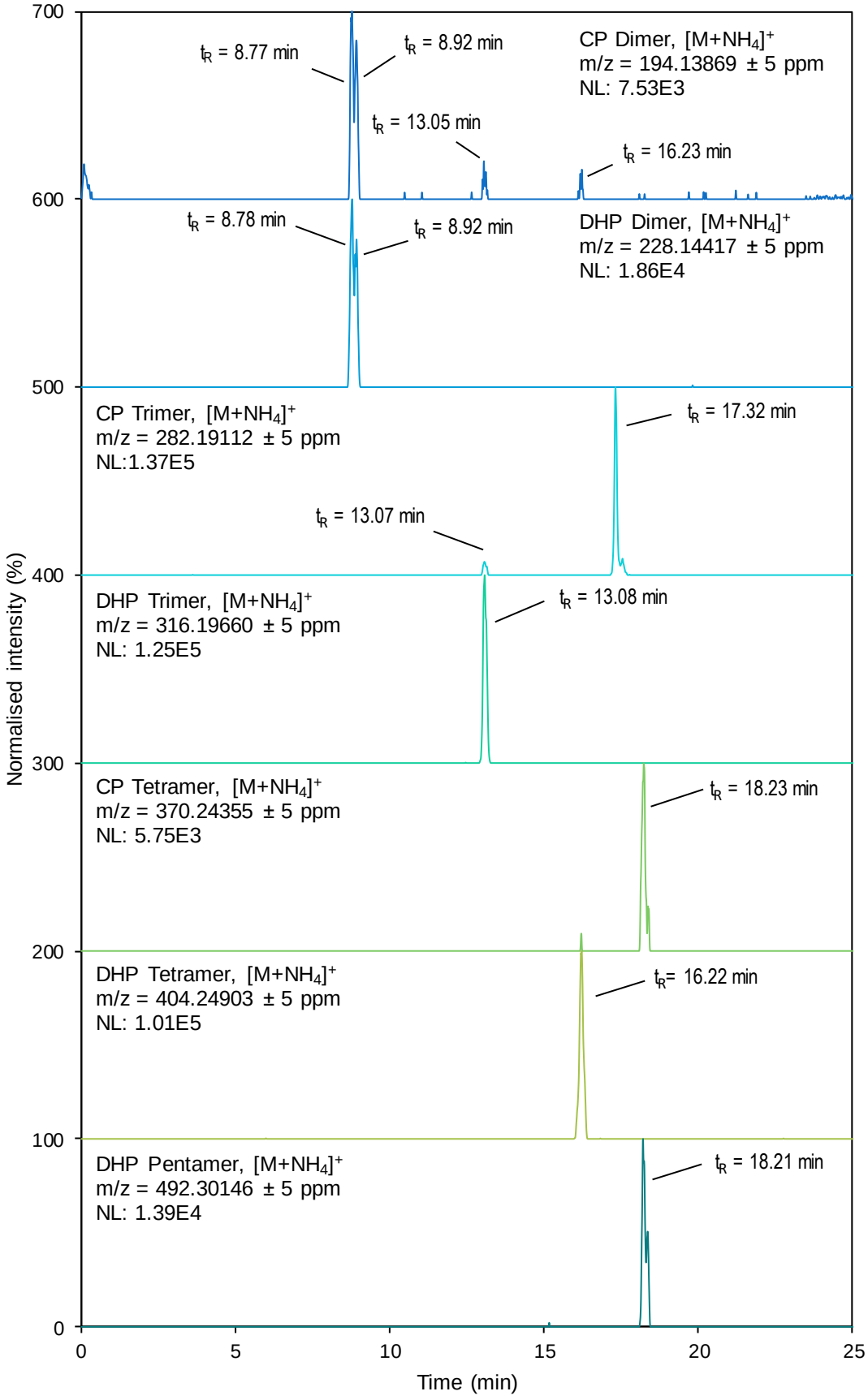


Figure 3.10: Extracted ion chromatograms showing the presence of multiple oligomers in a solution of MEKP.

first two peaks at 8.78 and 8.92 min were not fully resolved and both peaks had masses for the ammonium adducts of MEKP linear dihydroperoxy peroxide (DHP) and cyclic peroxide (CP) dimers, suggesting that they either interconverted at the ionisation source, or the DHP formed a $[M\text{-}H_2O_2\text{+}NH_4]^+$ fragment ion and the CP formed a $[M\text{+}H_2O_2\text{+}NH_4]^+$ adduct ion. The chromatographic peaks at 13.07 and 16.22 min were thought to be due to the DHP trimer and tetramer, respectively. Although ions with a *m/z* corresponding to the ammonium adducts of smaller cyclic peroxides were also detected at these retention times, it was thought that these were due to common fragment ions. The peak at 17.32 min was thought to be due to the MEKP CP trimer. The extracted ion chromatograms for the exact *m/z* ± 5 ppm of the ammonium adducts of both the MEKP DHP pentamer and the MEKP CP tetramer had a peak at 18.22 min which appeared to be split and so may have been due to two unresolved chromatographic peaks.

As individual standards were not available it was not possible to confirm the elution order of the MEKP oligomers. However, based on polarity it is logical that larger oligomers would be retained longer as they have a larger hydrophobic section which is able to interact with the non-polar C_{18} Ar stationary phase. It is also logical that DHPs would be less retained than CPs due to the polar hydroperoxy (-O-O-H) groups on the end of the DHPs which are able to hydrogen bond with the polar mobile phase. Retention time prediction, which will be discussed in the next chapter may help clarify the elution order. Collizza et al. also reported the detection of ammonium adducts of three MEKP CPs and four MEKP DHPs, when investigating acetonitrile ion suppression [115]. However, individual standards were not analysed and all analysis was performed by flow injection analysis, so no chromatographic separation was shown. DeTata et al. did also demonstrate chromatographic separation between the ammonium adducts for MEKP DHP trimer, tetramer, pentamer and hexamer [32], but detection of the MEKP CPs was not reported. It is unclear whether this was because their presence was not investigated or if different syntheses resulted in different mixtures of MEKP oligomers.

3.3.2.5 Sugars

Three ions were detected for the sugars (erythritol, mannitol, sorbitol and xylitol) in both positive and negative ionization mode. In positive ionization mode they were detected as the protonated molecule, ammonium adduct and a fragment ion (typically $[M\text{-}OH]^-$) but the negative ions were used for this study due to larger signal-to-noise ratios, especially for ions 1 and 2. The sugars and urea were unretained on the LC column and nitroguanidine (NQ) was barely retained. This is

unsurprising given the highly polar nature of these five compounds and the use of a non-polar, C_{18} Ar, stationary phase.

3.3.3 Instrumental method performance

For qualitative methods the most important performance criteria are selectivity and limit of detection. Selectivity will be discussed in greater detail in subsequent sections when considering the use of different identification criteria. Average retention time (t_R), relative standard deviation (RSD) of t_R and estimated limits of detection are shown in Table 3.3. for all three ions. When determining LODs it was necessary to first consider what constituted a peak. At higher concentrations, peak identification is relatively straight forward and unambiguous; but as the concentration is lowered a grey area is reached where different peak detection criteria lead to differing results. This was particularly the case here when using HRMS analysis, as background noise was not present in all extracted ion chromatograms and so it was not always possible to use a 3:1 signal-to-noise ratio to determine the limit of detection. However, use of standard deviation of response and slope to calculate LODs [108] provided an acceptable approximation. Since LODs were calculated from a slope, linear range and linearity (R^2) are also given in Table 3.3 and it was only possible to calculate LODs for ions detected in enough concentrations to plot a calibration line. This was not the case for any of the ions of 3-NT, NB or NM. Although the *m/z* range was set from 60.00-625.00 when the LOD experiment was performed, the lowest *m/z* actually detected was 60.0172, greater than the *m/z* of deprotonated NM (*m/z* 60.0091). For the detection probability experiment, and all future experiments, the *m/z* range was set from 59.90-625.00 enabling detection of deprotonated NM. LODs were at the low pg level on column for many ions. Interestingly, the lowest LOD of the three ions did not always correspond to the ion with the greatest signal-to-noise in a 10 ng μL^{-1} standard, as some calibration curves levelled off before reaching 10 ng μL^{-1}. Measured detection probability at a threshold level is arguably of more value than estimated limits of detection for forensic analysis. Here detection probability was calculated for ion 1 at 1 ng on column. Whether or not a compound was detected depended on the peak detection criteria used, including minimum peak width, the noise method and the amount of smoothing. It would also depend on identification criteria such as the number of ions required.

Table 3.3: LC-APCI-HRMS performance: estimated limits of detection (LOD), linearity, repeatability, reproducibility and detection probability (n= 61, MEKP oligomers not included)

Analyte	Av. t_R (min)[a]	RSD (%)[a]	Estimated LOD Ion 1 (pg)	Estimated LOD Ion 2 (pg)	Estimated LOD Ion 3 (pg)	Linear range (pg μL^{-1})[b]	Linearity (R^2)[b]	Peak area repeatability RSD (%)[b, c]	Peak area reproducibility RSD (%)[a, b]	Detection probability (%)[b, d]
1,2-DAP	3.90	0.0	N.D	N.A.	N.A.	N.D.	N.D.	N.D.	N.D.	0
1,2-DNB	8.12	1.1	19	13	95	1-1000	0.994	12	10	100
1,2-DNG	4.47	0.9	56	583	985	10-1000	0.996	12	19	100
1,3-DNB	9.18	1.0	18	84	194	1-1000	0.996	9	12	100
1,3-DNG	4.04	0.8	169	120	829	250-1000	0.998	8	16	100
2,4-DA-6-NT	3.59	1.6	5	155	1851	1-1000	0.997	11	8	100
2,4-DNT	11.44	0.9	36	N.D.	142	1-1000	0.998	10	16	100
2,6-DA-4-NT	3.15	1.4	5	150	1339	1-1000	0.996	10	8	100
2,6-DNT	11.27	0.9	24	24	N.D.	1-1000	0.993	12	14	100
2-A-4,6-DNT	10.88	1.0	20	12	527	1-1000	0.997	10	14	100
2-MNG	1.73	0.5	19	1199	N.D.	5-1000	0.996	13	14	100
2-NT	10.70	0.9	1977	N.D.	N.D.	250-1000	0.984	31	40	100
3,4-DNT	10.36	1.0	19	15	32	1-1000	0.993	11	12	100
3,5-DNA	9.42	1.1	30	3	N.A.	5-1000	0.992	13	14	-
3-NT	11.09	0.9	N.D	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	0
4-A-2,6-DNT	10.54	1.1	9	10	24	2.5-1000	0.994	10	16	100
4-NT	11.02	1.2	1033	N.D.	N.D.	500-1000	0.999	29	33	100
DDNP	4.14	1.2	605	690	N.D.	100-1000	0.843	18	23	100
DEDPU	15.08	0.6	10	N.D.	535	1-1000	0.995	12	10	100
DEGDN	6.85	0.9	8135	N.D.	2392	N.D.	N.D.	14	11	100
DMDPU	12.69	0.7	9	635	711	1-1000	0.991	13	10	100
DMNB	7.46	1.2	91	N.D.	9948	25-10000	0.994	14	14	30
DNSA	7.17	1.7	6	434	657	25-10000	0.981	19	17	100

Table 3.3 (Continued): LC-APCI-HRMS performance: estimated limits of detection (LOD), linearity, repeatability, reproducibility and detection probability (n= 61, MEKP oligomers not included)

Analyte	Av. t_R (min)[a]	RSD (%)[a]	Estimated LOD Ion 1 (pg)	Estimated LOD Ion 2 (pg)	Estimated LOD Ion 3 (pg)	Linear range (pg μL^{-1})[b]	Linearity (R^2)[b]	Peak area repeatability RSD (%)[b, c]	Peak area reproducibility RSD (%)[a, b]	Detection probability (%)[b, d]
DPA	14.24	0.7	8	N.A.	N.A.	1-1000	0.994	11	10	100
EGDN	5.52	0.8	1206	N.A.	N.A.	750-10000	0.995	22	21	100
ETN	12.02	0.7	1294	5183	N.D.	500-10000	0.997	18	19	100
HMDD	4.50	1.5	9	340	721	2.5-1000	0.997	9	13	-
HMTA	1.39	1.6	11	N.D.	628	2.5-1000	0.993	9	9	100
HMTD	2.84	19.9	101	49	474	25-10000	0.992	10	8	13
HMX	4.72	1.5	9	73	115	2.5-1000	0.986	18	18	100
HNDPA	17.57	0.7	21	17	25	1-1000	0.992	13	13	100
NB	7.75	1.1	N.D.	N.D.	N.D.	N.D.	N.D.	10	64	53
NG	9.16	0.8	697	612	1726	250-10000	0.996	19	18	93
NM	1.80	1.1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	40
NQ	1.52	1.4	6	2076	893	2.5-1000	0.995	6	8	100
PA	7.34	1.4	17	32	34	1-1000	0.992	14	14	100
PETN	12.67	0.7	659	361	1928	250-10000	0.998	6	12	100
PGDN	7.64	7.5	1476	N.D.	N.D.	100-10000	0.995	5	10	100
Picramic acid	7.61	0.8	6	30	25	2.5-1000	0.993	15	29	-
PYX	22.64	1.2	24	18	225	1-1000	0.988	7	23	100
RDX	6.05	0.9	15	101	147	2.5-1000	0.996	17	17	100
R-Salt	3.59	0.7	10	48	238	2.5-1000	0.992	14	16	100
TATB	9.52	0.7	19	414	209	1-1000	0.996	16	18	100
TATP	12.22	1.1	86	1048	655	25-10000	0.999	9	9	63
TEGDN	7.79	1.0	70	726	316	5-1000	0.995	11	14	100
Tetryl	12.07	0.7	26	36	1907	10-10000	0.993	12	11	100
TMETN	12.03	1.0	220	235	1879	50-10000	0.996	6	11	100
TNB	9.96	0.9	19	27	12	5-1000	0.995	11	11	100

Table 3.3 (Continued): LC-APCI-HRMS performance: estimated limits of detection (LOD), linearity, repeatability, reproducibility and detection probability (n= 61, MEKP oligomers not included)

Analyte	Av. t_R (min)[a]	RSD (%)[a]	Estimated LOD Ion 1 (pg)	Estimated LOD Ion 2 (pg)	Estimated LOD Ion 3 (pg)	Linear range (pg μL^{-1})[b]	Linearity (R^2)[b]	Peak area repeatability RSD (%)[b, c]	Peak area reproducibility RSD (%)[a, b]	Detection probability (%)[b, d]
TNT	12.48	0.0	24	17	16	1-1000	0.993	12	11	100
1-MNG	1.72[e]	0.6	-	-	-	-	-	-	-	-
Aspirin	1.45[f]	0.0	-	-	-	-	-	-	-	-
DADP	8.48[g]	1.0	-	-	-	-	-	-	-	-
Erythritol	1.46[f]	0.7	-	-	-	-	-	-	-	-
Mannitol	1.44[f]	0.7	-	-	-	-	-	-	-	-
MHN	14.51[g]	1.0	-	-	-	-	-	-	-	-
Picrate	7.61[h]	1.5	-	-	-	-	-	-	-	-
Salicylic acid	1.49[f]	1.0	-	-	-	-	-	-	-	-
SHN	14.50[g]	0.7	-	-	-	-	-	-	-	-
Sorbitol	1.44[f]	1.3	-	-	-	-	-	-	-	-
Urea	1.48[f]	0.7	-	-	-	-	-	-	-	-
Xylitol	1.44[f]	0.0	-	-	-	-	-	-	-	-

N.A. Not applicable as no Ion 2 and/or Ion 3 detected; N.D. Not determined as detected in insufficient concentrations/injections; "-" Analyte not included due to a lack of pure standard at time of experiment.

[a] Average, SD or RSD of n=18 injections over 3 days; [b] Calculated for Ion 1 only; [c] n=6 injections; [d] n=30 injections of 1 ng on column; [e] Not included in method performance mix as indistinguishable from 2-MNG; [f] Unretained and so no further LC-HRMS method performance assessed, n=2 injections; [g] Not included in method performance mix as no pure standard; [h] Not included in method performance mix as indistinguishable from PA;

3.3.4 Effect of using different identification criteria on selectivity

3.3.4.1 Selectivity achieved by HRMS

The accurate mass of an ion determined by HRMS can be used to determine the elemental composition. This is because, apart from carbon, elements do not have integer masses and therefore different elemental compositions result in different exact masses, even if they have the same nominal mass (i.e. isobars). However, the ability to separate isobars effectively and determine unique elemental compositions is dependent on high mass accuracy as well as high mass resolution. Here, the number of elemental compositions generated by instrument software was investigated with 5, 2 and 1 ppm mass accuracy thresholds.

This was a theoretical experiment and some caveats are required before interpreting the results. Firstly, elemental compositions were generated based on the monoisotopic mass of the molecule rather than the mass of an ion for simplicity. Not all compounds will form the same ions and other compounds may form adduct or fragment ions with a similar mass. However, without mass spectra it was difficult to predict which ions would for and this would likely differ between different instruments and ionisation sources. Currently, open access mass spectral databases are still limited, especially in comparison to the 63 million compounds included in ChemSpider [117]. For example, when this study was performed, MassBank contained 41,092 mass spectra but only 633 of those were acquired using LC-APCI-QTOF and none were acquired using an orbitrap mass analyser [118]. Additionally, elemental compositions were limited to the most abundant isotopes of the elements included in the ions detected in Table 3.2 (i.e. carbon, hydrogen, nitrogen, oxygen and chlorine) with the number of each element included limited only by the mass. Compounds containing other, less common, elements and/or isotopes may also have an elemental composition within the specified mass accuracy threshold. Conversely, compounds were not found (within the ChemSpider database at least) with all the theoretical elemental compositions generated. These restrictions were necessary to simplify the theoretical experiment and while the theoretical number of elemental compositions generated here will likely differ from the number of elemental compositions of known compounds or ions; since all compounds and mass accuracy thresholds were treated in the same manner the generated theoretical elemental compositions were still of value for assessing the effect of mass accuracy thresholds.

Another important consideration when identifying a compound from an exact mass is the number of isomers with the same elemental composition. For each of the explosives-related compounds

considered here, the number of compounds listed in the ChemSpider database with the same elemental composition (isomers) are shown in **Table** 3.4 along with the number of theoretical elemental compositions within 5, 2 and 1 ppm of the monoisotopic mass as generated by Xcalibur Qual browser. Ammonium picrate was not included in these theoretical experiments as it is not a single component structure and so would be excluded from the Chemspider search.

Table 3.4: Isomers and elemental compositions within 5, 2 and 1 ppm of monoisotopic mass (n=60 compounds, MEKP oligomers and ammonium picrate not included)

Analyte	Molecular formula	Monoisotopic mass	Number of isomers*	Elemental compositions† within 5 ppm	Elemental compositions† within 2 ppm	Elemental compositions† within 1 ppm
1,2-DAP	$C_3H_{10}N_2$	74.084396	16	1	1	1
1,2-DNB	$C_6H_4O_4N_2$	168.017105	54	2	1	1
1,2-DNG	$C_3H_6O_7N_2$	182.017502	7	2	2	2
1,3-DNB	$C_6H_4O_4N_2$	168.017105	54	2	1	1
1,3-DNG	$C_3H_6O_7N_2$	182.017502	7	2	2	2
1-MNG	$C_3H_7O_5N$	137.032425	9	2	2	2
2,4-DA-6-NT	$C_7H_9O_2N_3$	167.069473	1064	2	1	1
2,4-DNT	$C_7H_6O_4N_2$	182.032761	184	3	1	1
2,6-DA-4-NT	$C_7H_9O_2N_3$	167.069473	1064	2	1	1
2,6-DNT	$C_7H_6O_4N_2$	182.032761	184	3	1	1
2-A-4,6-DNT	$C_7H_7O_4N_3$	197.043655	184	3	1	1
2-MNG	$C_3H_7O_5N$	137.032425	9	2	2	2
2-NT	$C_7H_7O_2N$	137.047684	337	2	1	1
3,4-DNT	$C_7H_6O_4N_2$	182.032761	184	3	1	1
3,5-DNA	$C_6H_5O_4N_3$	183.028	72	3	1	1
3-NT	$C_7H_7O_2N$	137.047684	337	2	1	1
4-A-2,6-DNT	$C_7H_7O_4N_3$	197.043655	184	3	1	1
4-NT	$C_7H_7O_2N$	137.047684	337	2	1	1
Aspirin	$C_9H_8O_4$	180.042252	304	3	1	1
DADP	$C_6H_{12}O_4$	148.073563	396	1	1	1
DDNP	$C_6H_2O_5N_4$	210.002518	7	4	1	1
DEDPU	$C_{17}H_{20}ON_2$	268.157562	10083	4	3	1
DEGDN	$C_4H_8O_7N_2$	196.033157	5	3	2	2
DMDPU	$C_{15}H_{16}ON_2$	240.126266	4777	3	1	1
DMNB	$C_6H_{12}O_4N_2$	176.079712	265	1	1	1
DNSA	$C_7H_4O_7N_2$	228.001846	14	5	1	1
DPA	$C_{12}H_{11}N$	169.089142	164	1	1	1
EGDN	$C_2H_4O_6N_2$	152.006943	2	1	1	1
Erythritol	$C_4H_{10}O_4$	122.057907	14	1	1	1
ETN	$C_4H_6O_{12}N_4$	301.99823	3	19	9	5
HMDD	$C_6H_{12}O_4N_2$	176.079712	265	1	1	1
HMTA	$C_6H_{12}N_4$	140.106201	349	1	1	1

Table 3.4 (Continued): Isomers and elemental compositions within 5, 2 and 1 ppm of monoisotopic mass (n=60 compounds, MEKP oligomers and ammonium picrate not included)

Analyte	Molecular formula	Monoisotopic mass	Number of isomers*	Elemental compositions† within 5 ppm	Elemental compositions† within 2 ppm	Elemental compositions† within 1 ppm
HMTD	$C_6H_{12}O_6N_2$	208.069534	21	3	2	2
HMX	$C_4H_8O_8N_8$	296.046509	3	19	9	4
HNDPA	$C_{12}H_5O_{12}N_7$	438.999634	4	56	24	13
Mannitol	$C_6H_{14}O_6$	182.079041	38	3	2	2
MHN	$C_6H_8O_{18}N_6$	451.989502	4	69	30	12
NB	$C_6H_5O_2N$	123.032028	100	2	1	1
NG	$C_3H_5O_9N_3$	227.002579	2	4	1	1
NM	CH_3O_2N	61.016376	9	1	1	1
NQ	$CH_4O_2N_4$	104.033424	4	1	1	1
PA	$C_6H_3O_7N_3$	228.997101	7	4	1	1
PETN	$C_5H_8O_{12}N_4$	316.013885	1	22	10	5
PGDN	$C_3H_6O_6N_2$	166.022583	6	2	2	2
Picramic acid	$C_6H_5O_5N_3$	199.022919	39	3	1	1
PYX	$C_{17}H_7O_{16}N_{11}$	621.007202	3	223	91	46
RDX	$C_3H_6O_6N_6$	222.034882	6	4	2	2
R-Salt	$C_3H_6O_3N_6$	174.05014	6	2	1	1
Salicylic acid	$C_7H_6O_3$	138.031693	95	2	1	1
SHN	$C_6H_8O_{18}N_6$	451.989502	4	69	30	12
Sorbitol	$C_6H_{14}O_6$	182.079041	38	3	2	2
TATB	$C_6H_6O_6N_6$	258.034882	10	7	6	2
TATP	$C_9H_{18}O_6$	222.110336	112	3	2	2
TEGDN	$C_6H_{12}O_8N_2$	240.059372	1	4	3	2
Tetryl	$C_7H_5O_8N_5$	287.013824	4	15	6	2
TMETN	$C_5H_9O_9N_3$	255.033875	2	7	4	2
TNB	$C_6H_3O_6N_3$	213.002182	9	4	1	1
TNT	$C_7H_5O_6N_3$	227.017838	26	4	1	1
Urea	CH_4N_2O	60.032364	11	1	1	1
Xylitol	$C_5H_{12}O_5$	152.068466	16	2	2	2

* Single-Component structures in Chemspider database matching the molecular formula (Accurate as of 14/06/17)
† Theoretical elemental compositions generated by Xcalibur ($C_{(0-52)}H_{(0-620)}N_{(0-44)}O_{(0-39)}Cl_{(0-17)}$)

3.3.4.1.1 Effect of mass accuracy on number of possible elemental compositions

For the majority of compounds (n=49) more than one theoretical elemental composition fell within 5 ppm of the monoisotopic mass, as shown in Table 3.4, Figure 3.11 and Figure 3.12. This number was reduced when a stricter mass accuracy threshold was used and the median number of elemental compositions with a mass accuracy threshold of either 1 or 2 ppm was 1 (Figure 3.12). However, the absolute number of theoretical elemental compositions for each compound is of more interest than averages in forensic science and even with a 1 ppm mass accuracy

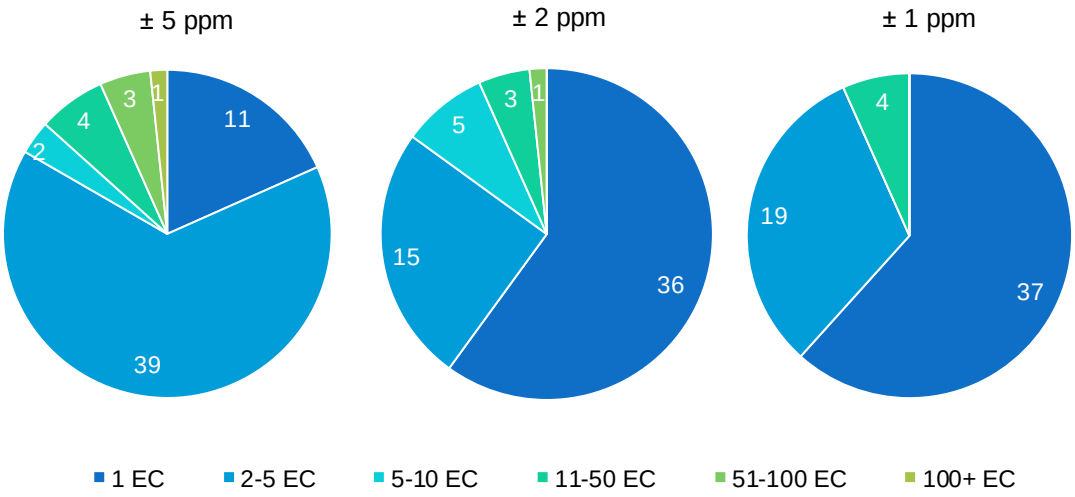


Figure 3.11: Pie charts summarising the number of elemental compositions (EC) within 5, 2 or 1 ppm of the exact mass of each compound (n=60, MEKP oligomers and ammonium picrate not included).

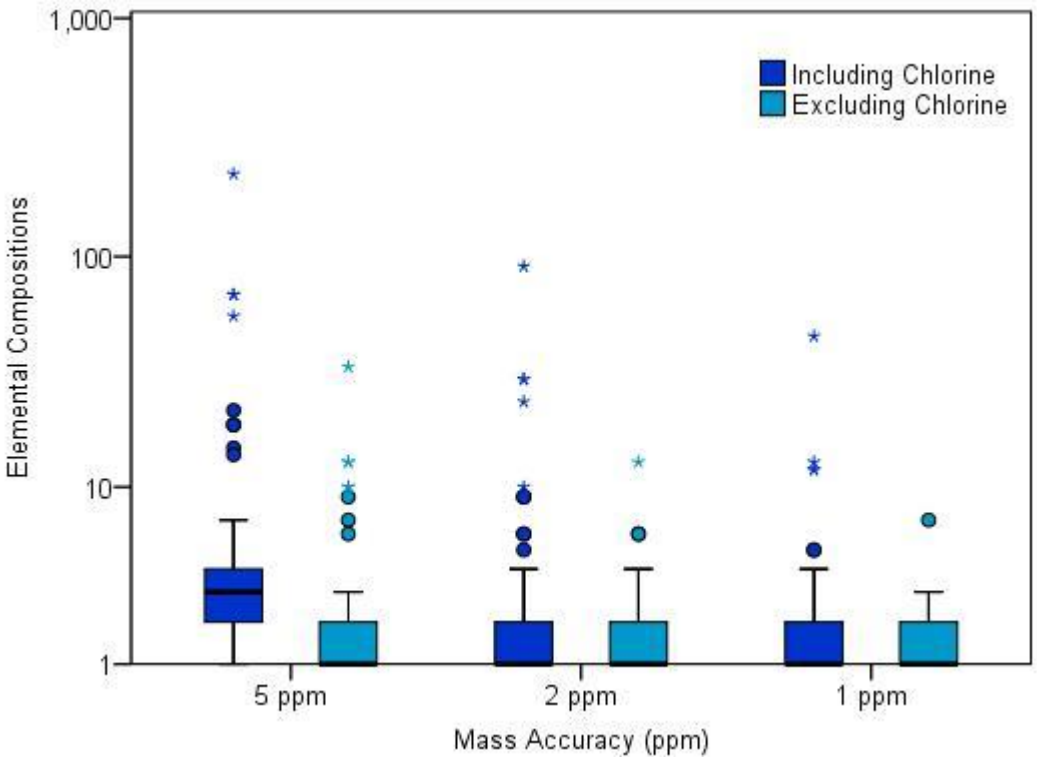


Figure 3.12: Boxplot showing the effect of mass accuracy thresholds and including/excluding chlorine on the number of elemental compositions (n=60, MEKP oligomers and ammonium picrate not included). Boxes indicate interquartile range, whiskers are 1.5 times the height of the box, circles show outliers with values 1.5-3 times height of box and stars show extreme outliers with values >3 times the height of the box.

threshold, 23 compounds had more than one theoretical elemental composition (Figure 3.11). As shown in the boxplot (Figure 3.12) the number of possible elemental compositions was much higher for outliers with a maximum of 223 possible elemental compositions within 5 ppm, 91 within 2 ppm and 46 within 1 ppm for PYX. This was reduced to 34 within 5 ppm, 13 within 2 ppm and 7 within 1 ppm when chlorine was excluded from possible elemental compositions. For this experiment, the inclusion of chlorine increased the theoretical number of element compositions

within each mass accuracy threshold. However, the relatively large mass defect of chlorine (-0.03115 Da) has previously been reported to increase HRMS selectivity when considering compounds in a National Institute of Standards and Technology (NIST) library [94]. Therefore, further work is required to more thoroughly investigate the effect of detecting chloride adducts on HRMS selectivity.

Unsurprisingly, positive correlation was seen between monoisotopic mass and the number of theoretical elemental compositions that fell within 1, 2 or 5 ppm of the exact mass of each compound (Figure 3.13). The data fit well to a cubic equation ($R^2 > 0.99$ for 2 and 5 ppm), although more compounds with a monoisotopic mass of > 350 Da would be required to fully characterize the relationship between mass and the number of theoretical elemental compositions. Figure 3.13 also showed that reducing the mass accuracy window from 5 to 2 or even 1 ppm had a greater effect on the number of elemental compositions for larger compounds (especially those with monoisotopic mass >400 Da) than smaller compounds. However, while theoretically the number of possible elemental compositions increases continuously with mass, this is not reflected by known compounds as demonstrated by Little et al. who showed that the number of compounds in the ChemSpider database peaked in the mass range of 300-400 Da and dramatically decreased after 600 Da [119].

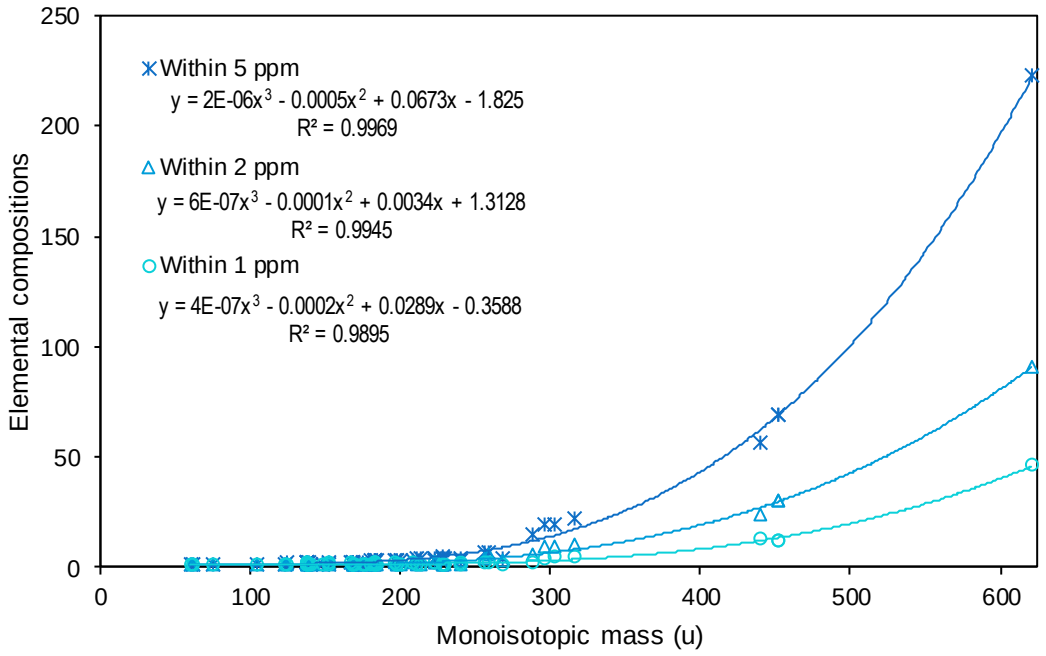


Figure 3.13: Correlation between monoisotopic mass and number of molecular formulae within 1, 2 and 5 ppm (n=60, MEKP oligomers and ammonium picrate not included)

When selecting a mass accuracy threshold, it was also important to consider the mass accuracy of the analytical method. The mass accuracy of ions detected for the compounds used in this study are shown in Figure 3.14. While over 90 % of ions had a mass accuracy within 2 ppm of the exact mass, the measured mass accuracy of this LC-HRMS method ranged from -2.5 to 3.8 ppm in standard solutions (Figure 3.14). Therefore, a mass accuracy window of 5 ppm was selected to reduce the risk of false negative results. While ultimately false positives are more problematic in Forensic Science, as they may lead to miscarriages of justice or reduced confidence in results; for a screening method, minimising false negatives is more important than minimising false positives. This is because false positives should be identified by subsequent confirmatory analysis but normally no further analysis would be performed following a negative result from a screen.

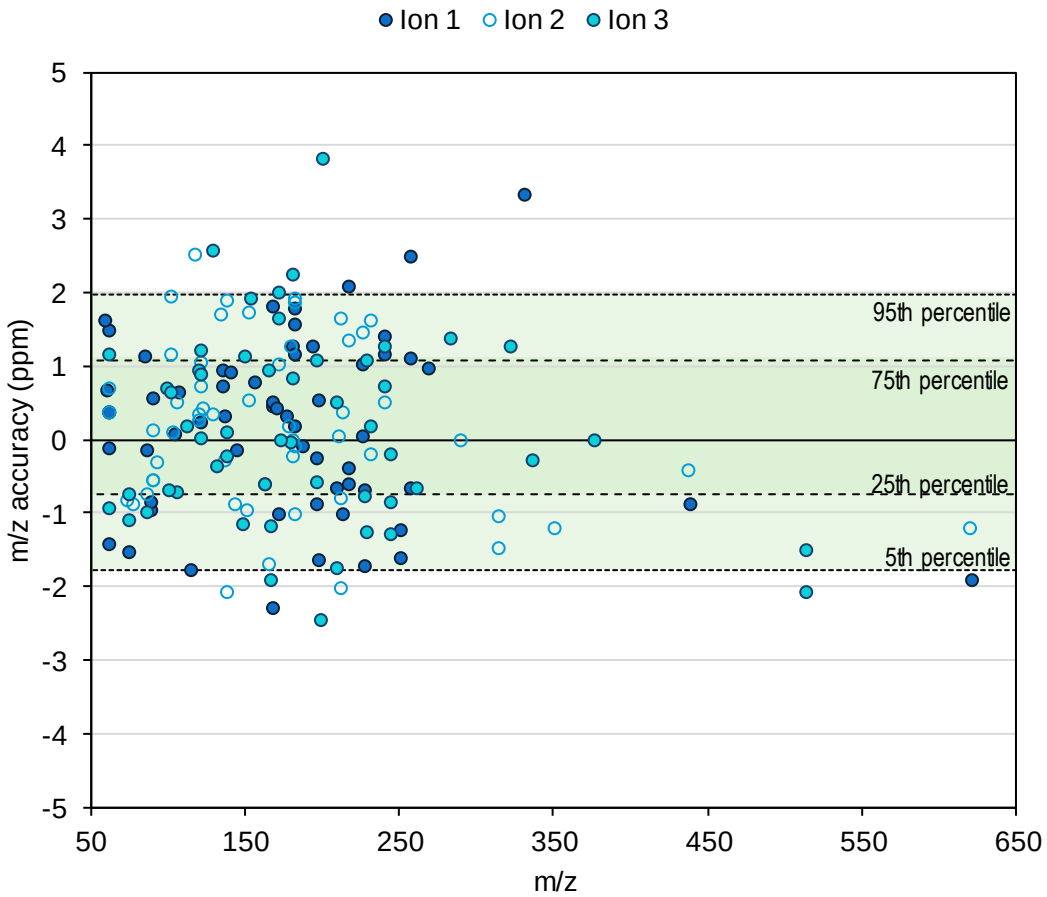


Figure 3.14: Mass accuracy of the three ions detected for each compound in a 10 ng μL^{-1} standard (n=61, MEKP oligomers not included).

With newer and improved instrumentation, it may be possible to apply stricter mass accuracy criteria. This could be worth considering in the future, especially for larger compounds where mass accuracy had a greater effect on the number of possible elemental compositions. However, even with improved mass accuracy it would be important to consider the possibility of multiple

elemental compositions since here it was not always possible to unequivocally determine the elemental composition even with a mass accuracy of 1 ppm.

3.3.4.1.2 Isotope patterns

Where there were multiple possible elemental compositions within the required mass accuracy threshold, isotope patterns could have been used to reduce the number of possible compositions. For example, if there was a chlorine included in the molecule or ion then the isotope pattern should include two peaks approximately two mass units apart with the first peak, due to the chlorine-35 isotope, approximately three times the intensity of the second peak, due to the chlorine-37 isotope, based on the natural abundancies. While chorine isotopes were relatively easy to identify visually for the chloride adducts, software is available that considers more complicated isotope patterns. One such software, TraceFinder 3.1 (Thermo Scientific) was briefly investigated. However, as TraceFinder 3.1 was unable to detect all compounds detected by Xcalibur in a reference standard, it was not pursued further. Negative mode molecular ions, $[M]^-$, were particularly problematic for TraceFinder 3.1 which automatically switched to the deprotonated molecular ion, $[M-H]^-$, when calculating theoretical isotope patterns for these ions. More recent versions of TraceFinder or alternative software may perform better and could be worth investigating in future studies to streamline data analysis. Regardless, it is also worth bearing in mind that rarer isotopes are unlikely to be detected close to the limit of detection for the most abundant isotope. Additionally, isotope patterns cannot distinguish between isomers since they have the same elemental composition.

3.3.4.1.3 Number of isomers with the same molecular formula

Figure 3.15 and Table 3.4 show the number of matches to a molecular formula search of the ChemSpider database. In almost all cases, there were multiple isomers with the same molecular formula. Pentaerythritol tetranitrate (PETN) and triethylene glycol dinitrate (TEGDN) were the only exceptions, This meant that even when the molecular formula could be confirmed by HRMS, it was not always possible to confirm the compound identity. In the most extreme case, diethyldiphenylurea (DEDPU), also known as ethyl centralite, there were 10,083 compounds listed with the same molecular formula. This was even after filtering for single component compounds only and disregarding isotopically labelled structures. Most compounds fell somewhere in between these two extremes.

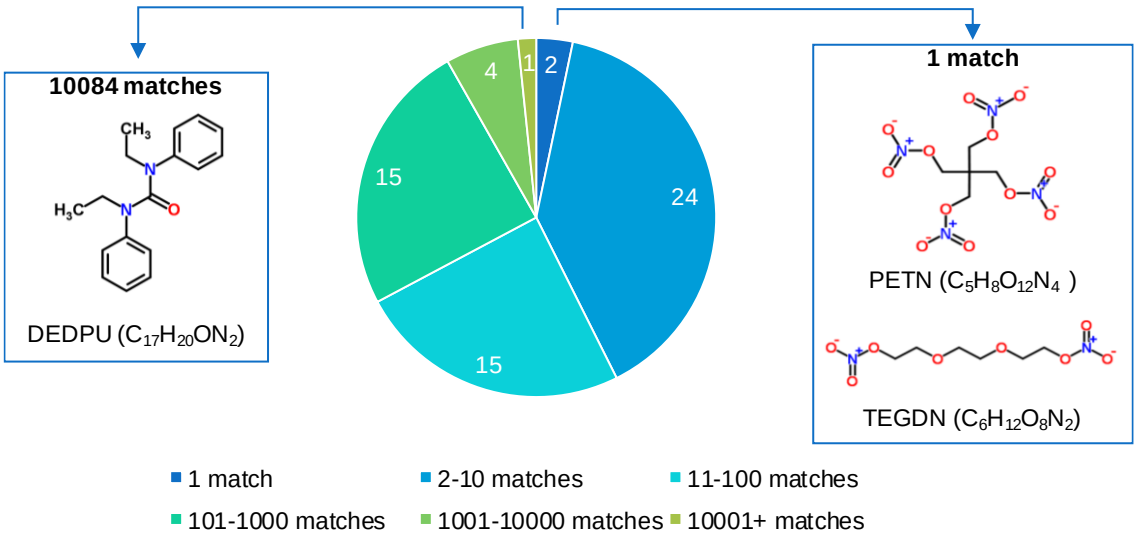


Figure 3.15: Pie chart summarising the number of matches from a molecular formula search of the ChemSpider database for each compound (n=60, MEKP oligomers and ammonium picrate not included).

It is worth considering the number of isomers when setting identification criteria. Often it is possible to separate pairs of isomers based on their chromatographic retention time, and/or fragment ions but the chance of other isomers having the same retention time and/or fragment ions will increase with the number of isomers. However, it would be extremely time consuming and costly to determine experimentally whether all 10,084 isomers of DEDPU could be distinguished based on their ions and/or retention times. This study has instead focussed on the more feasible aim of assessing selectivity both between 60 explosives-related compounds and from relevant matrix samples.

3.3.4.1.4 Selectivity between explosive compounds using HRMS data only

Due to structural similarities, many of the explosives-related compounds analysed here produced common ions. Therefore, it was important to investigate the ability to distinguish between target analytes (n=60) when assessing the selectivity of the method and setting identification requirements. Initially, when investigating which factors had the greatest influence on selectivity, the accurate masses detected were considered in isolation. Figure 3.16 shows the number of other target compounds (out of a maximum of 59) that also produced ions within 5 ppm of the exact mass of the ions identified in Table 3.2. In some cases, there were up to 15 other compounds that produced the same ion and could lead to a false positive if the ion with the greatest signal-to-noise (Ion 1) was the only identification criteria used. Adding the ion with the second greatest signal-to-noise (Ion 2) to the identification criteria reduced the number of false

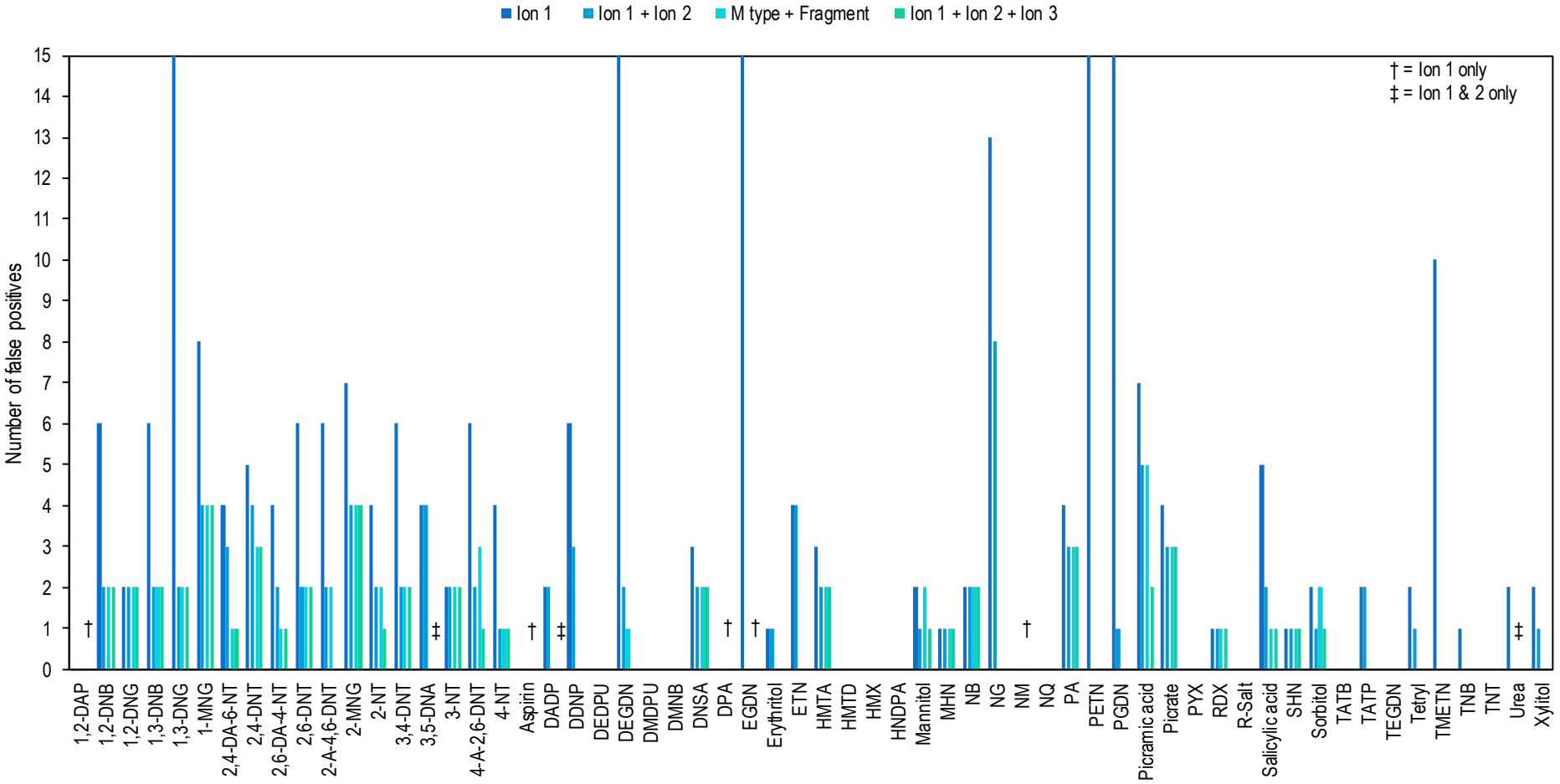


Figure 3.16: Selectivity achieved between explosive compounds when HRMS ions were used for identification without retention time windows (n=60, MEKP oligomers and HMDD not included). “M type” = molecular ion, (de)protonated molecule or adduct ion.

positives, but as not all compounds produced two ions it also led to false negatives in the case of acetyl salicylic acid (Aspirin), diphenylamine (DPA), ethylene glycol dinitrate (EGDN) and nitromethane (NM). In most cases, when the two ions had to be a molecular ion/(de)protonated molecule/adduct ion and a fragment ion (M type + Fragment) the number of false positives was reduced further. However, it is worth noting that this was not always the case and for 4-A-2,6-DNT, mannitol and sorbitol the two ions with the greatest signal-to-noise (Ion 1 and Ion 2) were more selective. Finally, in some cases (e.g. for DEGDN) a further reduction in false positives was achieved when all three ions were used. However, for 1,2-DNG, 3-NT, MHN, NB, RDX and SHN the number of ions used for identification had no effect on the number of false positives. Additionally, no false positives were seen for DEDPU, DMDPU, DMNB, DPA, HMTD, HMX, HNDPA, NM, NQ, PYX, R-Salt, TATB, TEGDN or TNT even when only Ion 1, the most abundant ion, was used as the identification criteria.

3.3.4.2 Selectivity achieved with a retention time window

The effect of including chromatographic retention time windows of average measured t_R ± 2.5 % in the identification criteria was examined for all compounds with a retention factor (k) greater than 1, in line with the chromatographic separation performance criteria outlined by the European Commission, 2002/657/EC [107]. As shown in Figure 3.17, including a retention time window in the identification criteria led to a greater reduction in the total number of false positives than the inclusion of additional ions. When using the ion with the greatest signal-to-noise ratio (Ion 1) and

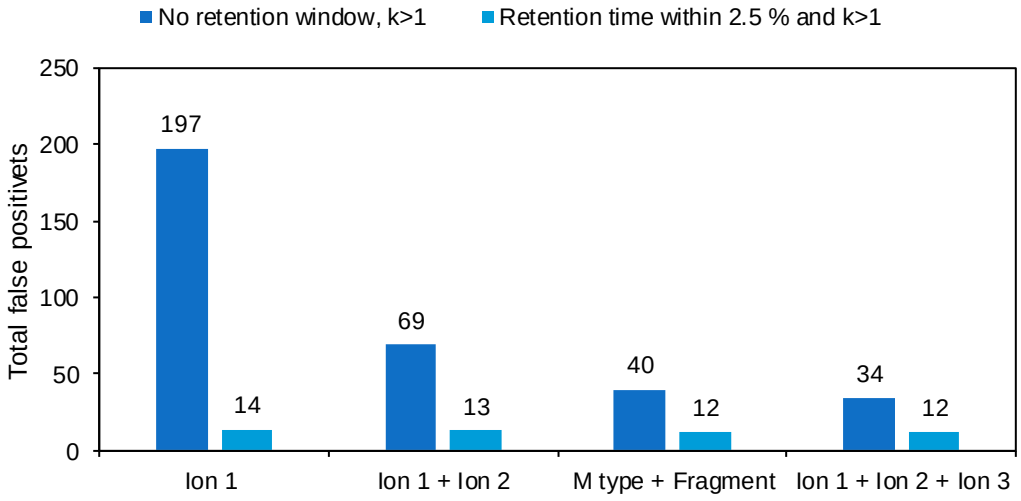


Figure 3.17: A comparison of the number of explosive related compounds resulting in false positives (within 5 ppm) with and without the use of retention time windows for identification of compounds with a retention factor (k) greater than 1 (n=47).

a retention time window of ± 2.5 %, only 14 false positives were detected, compared to 197 without any retention time windows. Erythritol tetranitrate (ETN) had a common fragment ion with trimethylolethane trinitrate (TMETN) and their retention times were within 2.5 %, but outside the 0.1 min tolerance recommended by SANTE/11813/2017 [107]. Picric acid and picrate from ammonium picrate appeared identical to this LC-HRMS method and both also shared three ions with dinitrosalicylic acid (DNSA) which also fell within the 2.5 % retention window. The remaining selectivity issues were due to pairs of isomers: 2- and 4- nitrotoluene (NT), mannitol and sorbitol hexanitrate, 2-amino and 4-amino-dinitrotoluene (A-DNT) and 2,4- and 2,6-dinitrotoluene (DNT). Stricter retention windows could separate some of these pairs, but also led to more false negatives. When using a retention window of ± 2.5 %, in most cases additional ions had no effect on the number of false positives. The only exceptions were for TMETN, where inclusion of any second ion provided sufficient selectivity from ETN and 2-A-4,6-DNT where inclusion of a fragment ion provided sufficient selectivity from 4-A-2,6-DNT.

3.3.4.3 Ion ratios for discriminating between isomers

Ion ratios were investigated to aid discrimination between isomers such as the dinitrotoluene (DNT) isomers (Figure 3.18). 2,4-DNT was the only isomer which favoured formation of the deprotonated molecule rather than the molecular ion (Figure 3.18). A one-way multivariate analysis of variance (MANOVA) revealed that there was a statistically significant difference ($p<0.0005$) in all three ion ratios obtained based on the different DNT isomers. Tukey's post-hoc test revealed a significant difference between 2,4-DNT and the other two isomers for all three ion ratios ($p<0.0005$). No significant difference was found between 2,6- and 3,4-DNT for the $[M-H]^-/[M]^-$ and $[M-NO]^-/[M]^-$ ion ratios ($p=0.146$ and 0.032 respectively) but a significant difference was seen between these isomers for the $[M-OH]^-/[M-H]^-$ ion ratio ($p<0.0005$). It is worth noting that $[M-OH]^-$ was not one of the three most abundant ions for either 2,6- or 3,4-DNT and so using this ion as a detection criterion would adversely affect sensitivity. Given that these two isomers were easily separated by chromatography, the cost in sensitivity was not outweighed by an improvement in selectivity. For 2,4- and 2,6-DNT, on the other hand, ion ratios were of greater value as these isomers were not fully resolved chromatographically but could be distinguished based on the ion ratio of the two most abundant ions ($[M-H]^-/[M]^-$). All ion ratios shown in Figure 3.18 were determined using EIC peak height in standards and so further work may be required to investigate the effect of matrix on the reproducibility of these ratios.

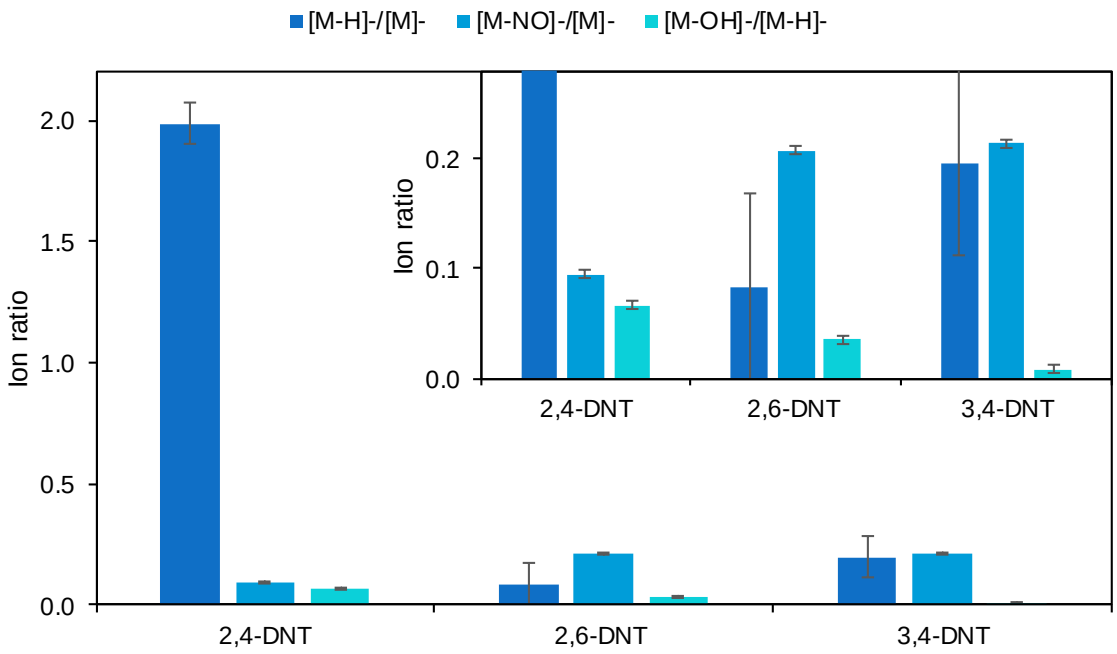


Figure 3.18: Ion ratios for 2,4-, 2,6- and 3,4-dinitrotoluene isomers. Error bars represent 95 % confidence intervals, n=6.

3.3.5 Selectivity in a selection of matrices

Selectivity from matrix components was also investigated using 18 samples of blank matrix, 6 natural fingermarks from different donors, 6 indoor dirt swabs and 6 outdoor dirt swabs. Figure 3.19 shows how many of the 18 matrix samples were positive for each compound when different HRMS ions were required for identification without any retention time restrictions. Figure 3.20 shows that the number of false positives reduced when chromatographic retention time was also included in the identification criteria. Fewer compounds were included in Figure 3.20 than Figure 3.19 because while it was possible to compare the ions detected for unretained compounds (e.g. the sugars), the lack of chromatographic retention prevented the use of retention time as an identification requirement. For the majority of retained compounds, no false positives were detected when a retention time window of ± 2.5 % was used. The only exceptions were 1,3-DNB, 2,4-DNT, 2,6-DNT and TNT, with 2,4-DNT the only compound for which multiple ions were detected. Examination of the negative controls used for these experiments showed no ions in the solvent blanks but the 2,4-DNT ions were also present in SPE extracts of blank cotton swabs, indicating that either the cotton swabs or the SPE cartridges were the source of these ions rather than the dirt collected on the swabs.

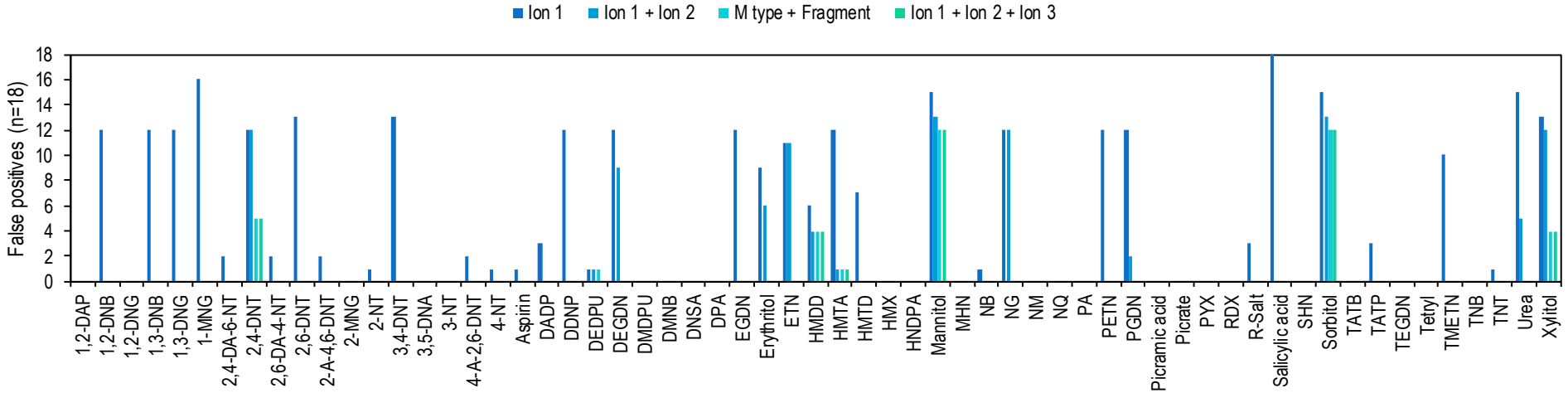


Figure 3.19: Selectivity issues in 18 matrices when relying on HRMS ions (*m/z* ± 5 ppm) only for identification of target compounds (n=61, MEKP oligomers not included).

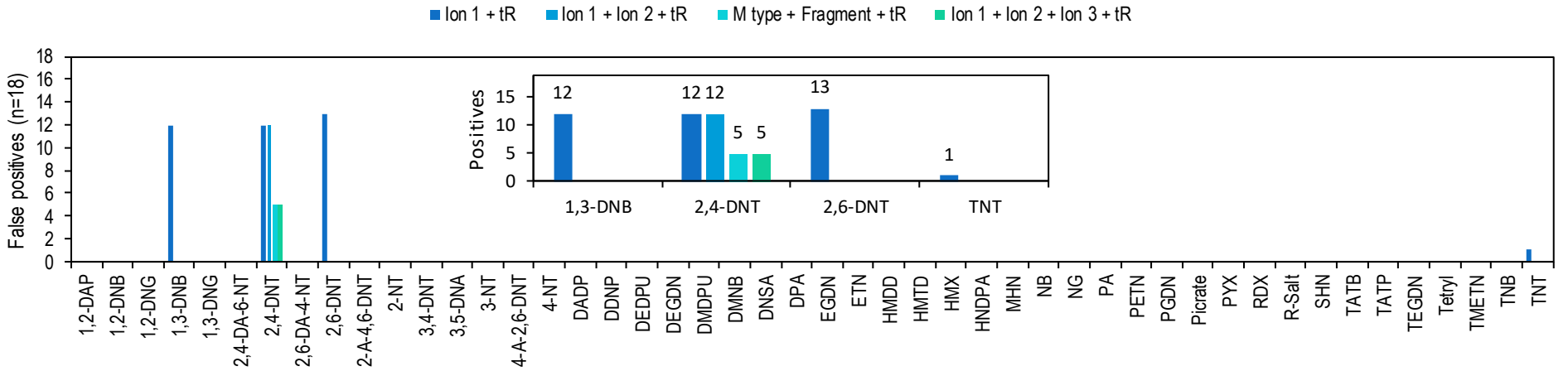


Figure 3.20: Selectivity issues in 18 matrices when both HRMS ions and retention windows (*m/z* ± 5 ppm, t_R ± 2.5 %, k > 1) were used for identification of target compounds (n=47).

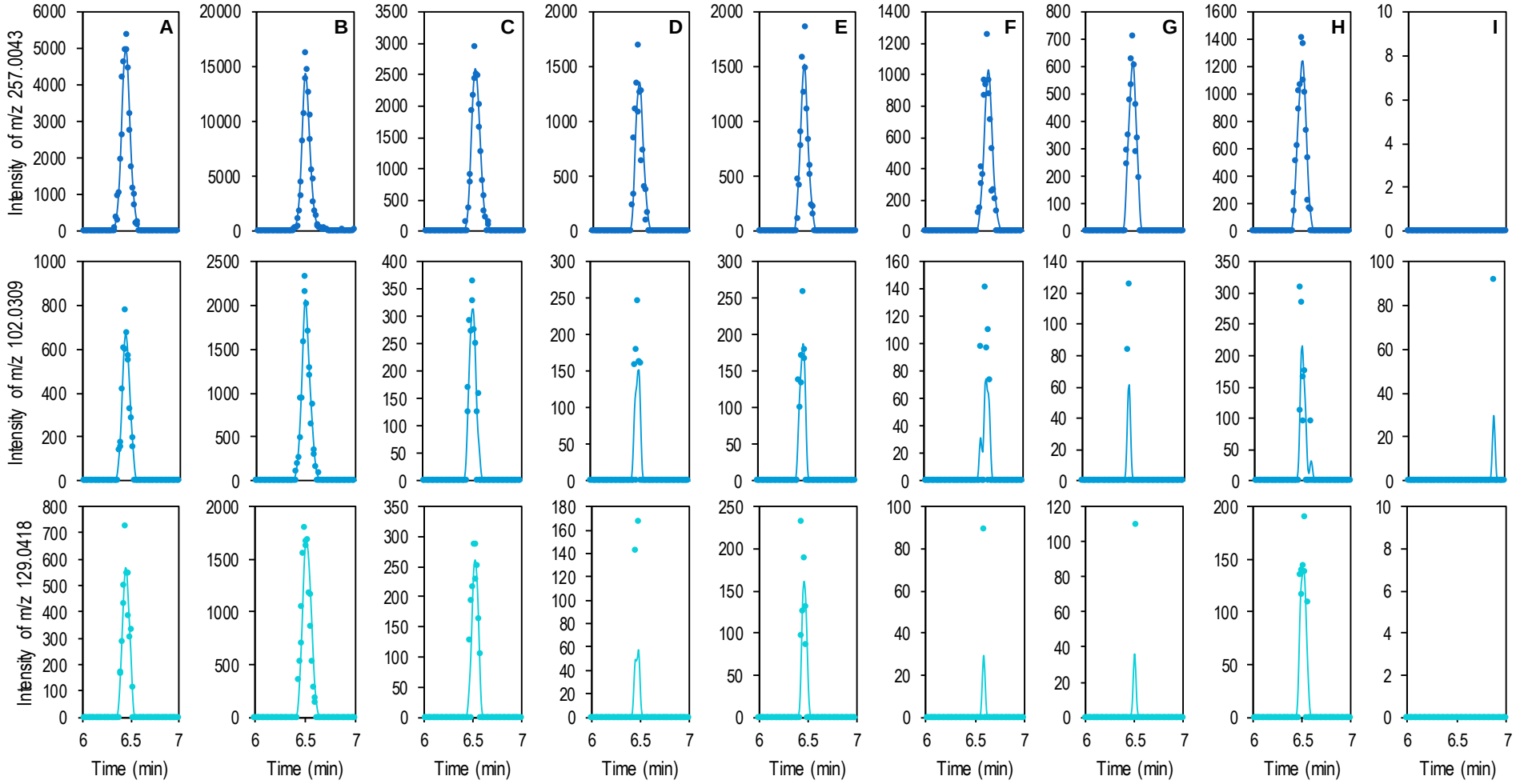


Figure 3.21: Fingermark depletion series EICs for RDX ion 1 (257.0043 ± 5 ppm, dark blue), RDX ion 2 (102.0309 ± 5 ppm, mid blue) and RDX ion 3 (129.0418 ± 5 ppm, light blue). A 0.01 ng μL^{-1} RDX standard is shown in A; the 1st, 11th, 21st, 31st, 41st, 51st, and 61st fingermark of the depletion series are shown in B-H and I shows a blank fingermark. Lines show 7-point gaussian smoothing.

3.3.6 Detection of fragment ions in fingermark depletion series

Having identified the presence of multiple ions for RDX and HMX it was decided to revisit the fingermark depletion series analysed in Chapter 2. The extracted ion chromatograms for RDX ions 1, 2 and 3 are shown in Figure 3.21. Using the Xcalibur Qual Browser default peak detection settings, without smoothing, the two fragment ions were only detected in the first fingermark, although changes in intensity were also seen in later fingermarks. Even the 0.01 ng uL^{-1} standard did not meet the default peak detection requirements. This highlights the compromise in sensitivity when multiple ions are included in the detection requirements.

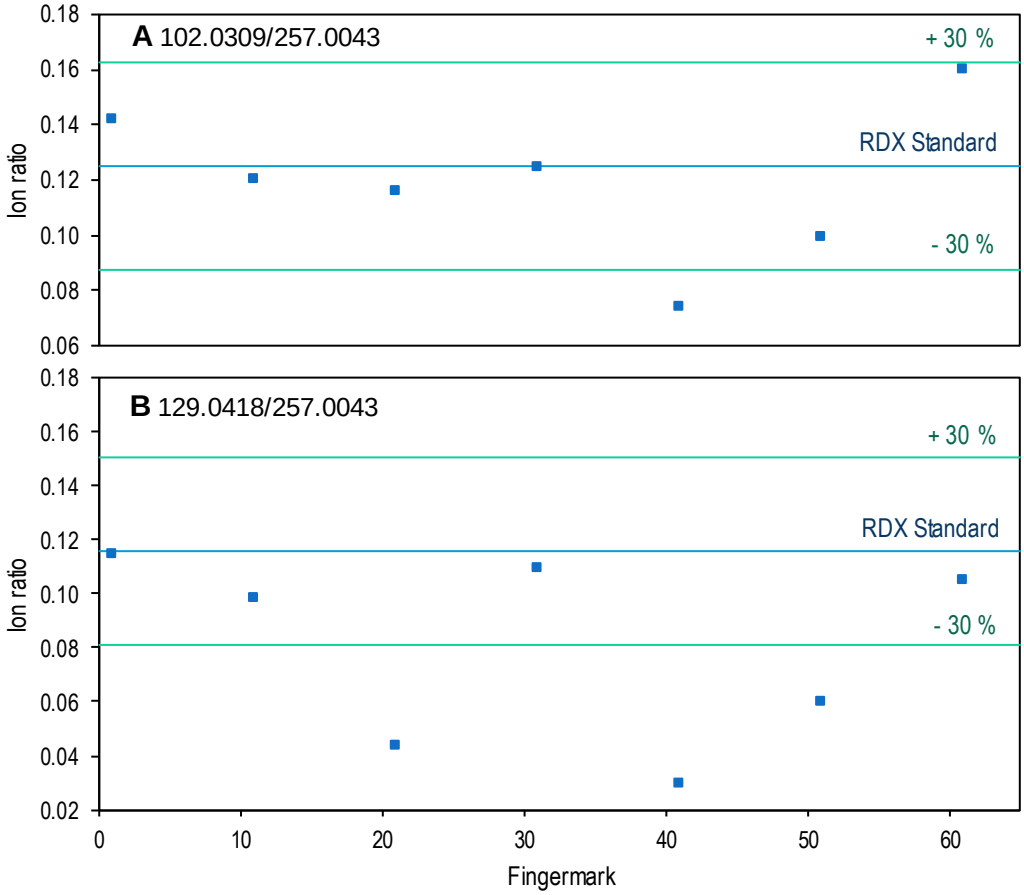


Figure 3.22: Ion ratios for RDX fragments versus chloride adduct ions in fingermark depletion series. Lines show average ion ratio in a 1 ng μL^{-1} RDX standard (n=6) and ± 30 %, the permitted tolerance for ions with relative intensity of 10-20 % according to 2002/657/EC [93].

Ion ratios were also calculated, based on the peak height of the smoothed EIC of the fragment ions relative to the chloride adduct (Figure 3.22). For six of the seven fingermarks analysed from the depletion series, the 102.0309/257.0043 ion ratio fell within ± 30 % of the average ion ratio in a 1 ng μL^{-1} RDX standard, the permitted tolerance according to 2002/657/EC [93]. In the case of the 129.0418/257.0043 ion ratio, four of the seven fingermarks fell within ± 30 % of the RDX standard. In all cases where ion ratios fell outside the permitted ± 30 % tolerance, fragment ion

'peaks' had to be manually integrated as even with smoothing they were not detected by the software. Therefore, these results support the use of ion ratios for in-source fragment ions, with the proviso that more concentrated samples are required to detect less abundant ions.

3.3.7 Method application to passive vapour sampling of explosives

Passive vapour samplers were screened for explosives following exposure to an explosives magazine for 24 hours. Extracted ion chromatograms for the ions detected are shown in Figure 3.23. While the explosive magazine was known to contain explosives, no information was provided regarding the type of explosives currently or previously stored in the magazine. Environmental conditions and dimensions of the building were also not known. It is acknowledged that this information would be necessary for replication of the experiment and should be recorded where possible. However, in this instance, even if the total volume of the magazine was known this would not correspond to the volume of air, since the magazine was not empty. Additionally, it was not a closed environment throughout the experiment as Explosive Technicians working on the site needed access to the magazine. Despite these limitations, 12 compounds were detected, demonstrating the potential of this screening method.

All three ions included in Table 3.2 were detected for 1,2-DNB, 1,3-DNB, 2,4-DNT, 3,4-DNT and TNT (Figure 3.23 A-C) within 2.5 % of the retention time of a standard; although the $[M-H]^-$ peak for 3,4-DNT had an intensity of <1 % of the 2,4-DNT peak for that ion. Two ions were also detected for 2,6-DNT but Ion 3, the $[M-H]^-$ ion, was missing in this case. This may have been due to the greater intensity of the 2,4-DNT peak, which was not fully resolved from 2,6-DNT, masking this peak. The dinitrobenzene (DNB) and dinitrotoluene (DNT) isomers are likely to have originated from TNT based explosives since 1,3-DNB, 2,4-DNT, 2,6-DNT and 3,4-DNT have all been detected as impurities in TNT previously [120]. TNT is often used in military explosives and comes in a range of forms including flake which can be melted and cast into specific shapes (e.g. to fill shells). The greater intensity of the 2,4-DNT peaks detected compared to TNT may be explained by the greater vapour pressure of 2,4-DNT (4.1×10^{-7} atm at 25 °C, compared to 9.2×10^{-9} atm for TNT [121]). Alternatively, smokeless powders containing 2,4-DNT may also have been present in the magazine.

Two ions were also detected for the nitrotoluene isomers, 2-NT and 4-NT, within 2.5 % of the retention time of a standard (Figure 3.23 D), although in both cases the second ion was close to the limit of detection. Therefore, the lack of a third ion may be due to this ion being below the limit

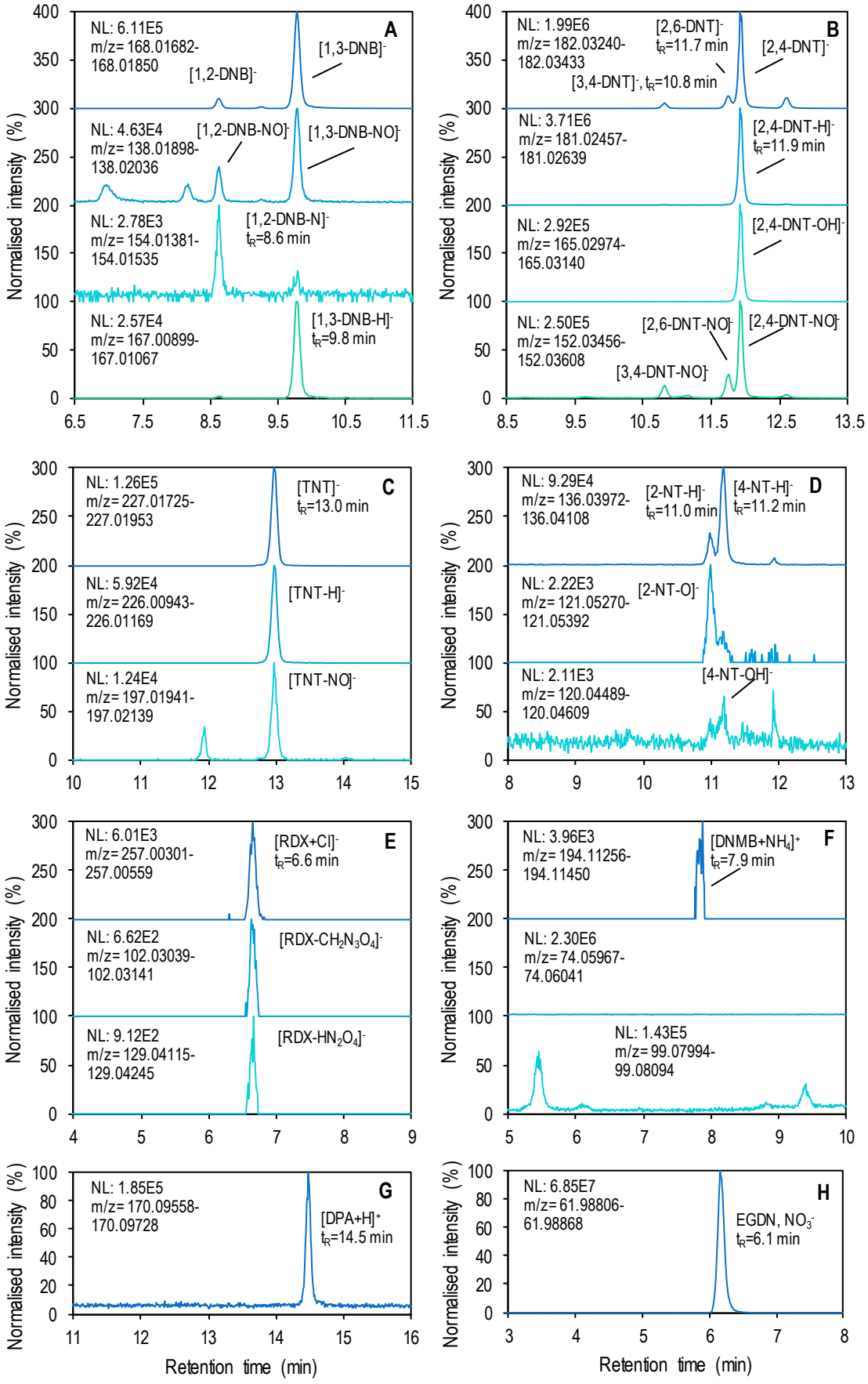


Figure 3.23: Passive vapour sampling of an explosives magazine with overlaid extracted ion chromatograms showing Ions 1-3 for: A) 1,2- and 1,3-DNB; B) 2,4-, 2,6- and 3,4-DNT; C) TNT; D) 2- and 4-NT; E) RDX; F) DMNB; G) DPA and H) EGDN.

of detection but, depending on the identification requirements used, 2- and 4-NT may be reported as not detected. The presence 2- and 4-NT vapours in the explosive magazine could have been due to these compounds being impurities in TNT explosives since they are intermediates in the synthesis of TNT, like the DNT isomers. Alternatively, they may have been used as detection agents to facilitate detection of plastic explosives via the vapour phase [122]; due to their high vapour pressure in comparison to PETN, RDX and HMX (~6.5 x 10^{-5} atm at 25 °C, compared to 1.1 x 10^{-11}, 4.9 x 10^{-12} and 2.4 x 10^{-17} atm for PETN, RDX and HMX, respectively [121]).

Changes in intensity were also seen in the extracted ion chromatograms for all three ions of RDX (Figure 3.23 E), within 2.5 % of the retention time of a standard, but intensities were very low and only two of the three ions were detected as peaks by the instrumental software without any smoothing. The presence of RDX in the magazine was unsurprising given that RDX is used in plastic explosives such as C4 and SEMTEX. However, given the low vapour pressure of RDX (2.4 x 10^{-17} atm at 25 °C [121]), detection of RDX on the passive vapour sampler was perhaps more surprising. However, Nacson and Grigoriev demonstrated that dust aided the detection of RDX in air [123]. Therefore, airborne particles containing RDX may have been detected here, rather than RDX vapours.

One ion was also detected for 2,3-dimethyl-2,3-dinitrobutane (DMNB), diphenylamine (DPA) and EGDN (Figure 3.23 F-H). In the case of DPA and EGDN only 1 ion was detected even in the highest concentration of standard and limits of detection were poor for the other two ions of DMNB which may explain the presence of only one ion. However, in all cases use of an alternative method would be recommended if confirmatory identification was required. DMNB and EGDN may also have been used as a detection agent to facilitate detection of plastic explosives via the vapour phase [122]. EGDN is also used in commercial blasting explosives which may explain why EGDN had the largest intensity of all the peaks detected. DPA is used as a stabiliser for nitrocellulose explosives and propellants [124], which could explain the presence of DPA in the magazine.

Ion ratios were calculated using EIC peak heights, for compounds where multiple ions were detected, to provided added confidence of correct identification. Figure 3.24 shows the calculated ion ratios for ions detected on three exposed passive samplers, compared to the average (n=6)

ion ratio of a standard in solution. The upper and lower threshold used, shown as green lines in Figure 3.24, was dependent on the relative abundance of the least abundant ion, with a threshold of ± 20, 25, 30 or 50 % used for relative abundancies > 50 %, 20-50 %, 10-20 % or < 10 % respectively [93]. As shown in Figure 3.24 A-C, for all three exposed passive samplers, all ion ratios calculated for ions assigned to 1,2-DNB (square marker, solid line) and 1,3-DNB (round marker, dotted line), based on retention time, fell within the relevant ion ratio thresholds. Additionally, there was no overlap between any of the ion ratio thresholds for 1,2- and 1,3-DNB, demonstrating the value of ion ratios for discriminating between isomers.

Results for the dinitrotoluenes (Figure 3.24 D-F): 3,4-DNT (square marker, solid line), 2,6-DNT (round marker, dotted line) and 2,4-DNT (diamond marker, dashed line) were more mixed. For all three passive samplers, the $[M-H]^-/[M]^-$ ion ratios for 2,4-DNT fell within the appropriate threshold, but the 3,4-DNT ion ratios fell below the lower threshold for 3,4-DNT but within the overlapping thresholds for 2,6-DNT (Figure 3.24 D). The $[M-NO]^-/[M]^-$ ion ratios (Figure 3.24 E) for all three DNTs fell within their respective thresholds and while the thresholds for 3,4-DNT and 2,6-DNT overlapped, neither overlapped with the thresholds for 2,4-DNT. The $[M-OH]^-/[M]^-$ ion ratios (Figure 3.24 F) for 2,4-DNT, also fell within the ion ratio thresholds for 2,4-DNT and outside the thresholds for 2,6- and 3,4-DNT. All calculated TNT ion ratios fell within the threshold levels (Figure 3.24 G and H). For RDX (Figure 3.24 I and J), the only calculated ion ratio that fell outside the ± 30 % threshold was for the third passive sampler where the intensity of RDX ions detected was lower than for the other two passive samplers. Overall, the use of ion ratios for the in-source fragment ions detected here showed promise, especially for the discrimination of isomers. Additionally, in most cases, the use of reference ion ratios determined in solution rather than matrix was not problematic. However, unlike the use of chromatographic retention time, using ion ratios for greater selectivity comes at a cost of poorer sensitivity as it is reliant on detection of a less abundant ion and so may result in more false negatives.

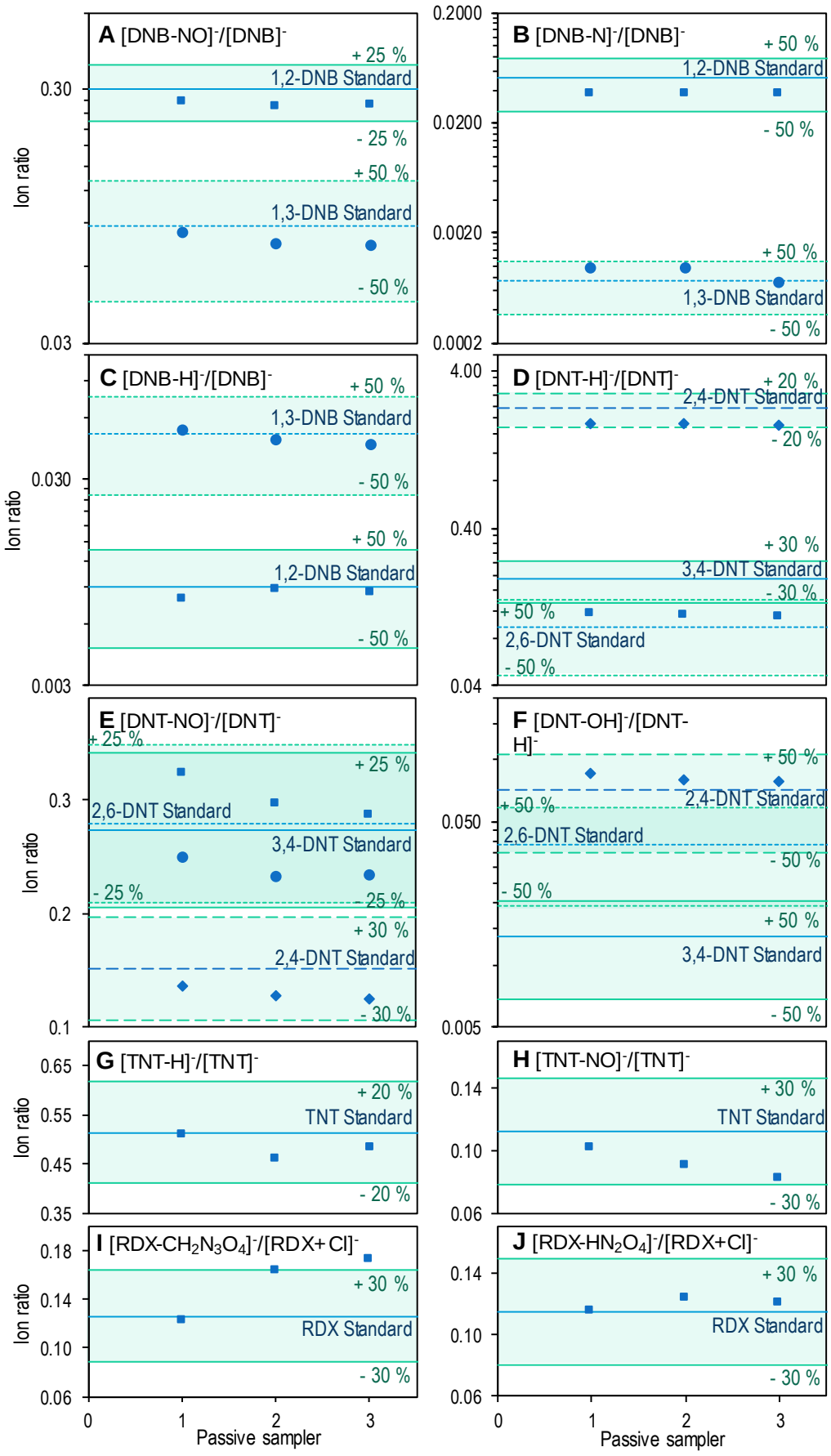


Figure 3.24: Ion ratios for ions detected on passive samplers. Blue lines show average ion ratio in standard (n=6) and green lines show upper and lower thresholds.

3.3.8 Alignment with method validation guidelines

In this chapter, the aim was to assess performance and understand the capability of this LC-HRMS method, without having a set of recognised criteria that the method must meet for forensic application. The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) defined validation as the "provision of objective evidence that a given item fulfils specified requirements, where the specified requirements are adequate for an intended use" in the ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories [125]. Method validation, against a set of specified requirements or performance criteria, should be performed before implementation in a forensic laboratory, and the analysis of casework samples, to demonstrate fitness for purpose. No specific validation guidelines were found for the field of forensic explosives analysis. However, the European Network of Forensic Science Institutes (ENFSI) recommended following the International Union of Pure and Applied Chemistry (IUPAC) and Eurachem validation guidelines [126-128].

For qualitative method validation, ENFSI highlighted the following performance parameters: precision; trueness or bias, in terms of the false positive and negative rates; measurement range, specifically a LOD or threshold and ruggedness, the effect of variation in individual test parameters on the outcome of the test [126]. The rate of false positives is also a measure of the selectivity of a method which was considered in Sections 3.3.4 and 3.3.5. Limits of detection were also estimated (Table 3.3). Robustness is another term for ruggedness and both refer to how reproducible results are with small changes [129]. Reproducibility of retention times and peak areas over three different days was investigated in this study (Table 3.3). Other variables that might affect ruggedness, such as different analysts, different lots of reagents and different environmental temperatures may also need to be assessed in the future However, as the ENFSI guidelines were very general, specific performance criteria and experimental design requirements for validation of LC-HRMS methods were not provided.

Both the Eurachem and the ICH validation guidelines stated that selectivity/specificity were the only performance criteria requiring validation for identification (qualitative) methods [108,128]. Presumably, this refers to identification of the main component of a sample which might be expected to be far above the LOD. For trace detection of explosives, LOD would also be important for interpretation as a negative result could be due to levels being below the LOD, rather than not

present at all. The Eurachem and ICH guidelines both recognised the importance of LOD as well as selectivity/specificity for limit testing of impurities methods [108,128], which would be more similar to trace detection of explosives. The IUPAC guidelines did not specify which performance characteristics are relevant to qualitative methods [127].

Whilst method validation against a set of specified requirements has not been carried out here, both LOD and selectivity have been evaluated approximately. Here an instrumental LOD was estimated based on 3.3 times the standard deviation of the response of 6 replicates divided by the slope of a calibration line (**Error! Reference source not found.**), one of the methods outlined for calculating LOD in the ICH guidelines [108]. The IUPAC guidelines highlight some of the challenges with establishing an accurate detection limit and the value of using a simple definition to provide a rough estimation of LOD whilst also recognising that instrumental LODs are often too low and inappropriate for method validation [127]. Therefore, the IUPAC guidelines recommend estimating precision based on the standard deviation (S_0) in 6 or more matrix blanks or low-level materials and using $3S_0$ to calculate an estimated LOD [127]. The Eurachem guidelines also recommend the use of three times standard deviation to estimate LOD; but using S'_0 which factors in the number of replicates used to estimate standard deviation [128]. While 10 replicates are recommended in the Eurachem quick reference on LODs, the guidelines also stated that the number of replicates should be “sufficient to obtain an adequate estimate of the standard deviation”, which they say typically requires between 6 and 15 replicates [128]. The Eurachem guidelines do also recognise that not all laboratories have the time and resources to carry out all the validation experiments described and using fewer replicates will still yield useful information [128]. For forensic analysis it is important that any uncertainties relating to the methods used are well understood to avoid miscarriages of justice.

For suspect or non-target screening it is hard to see how methods could be validated in advance, without reference materials. Following a positive result, confirmation would likely be required using a validated method for results to stand up in court. However, when using a screening method, knowledge of the rate of false negatives is important for making decisions since a negative result from a screen will often result in no further analysis. According to the European Commission Decision 2002/657/EC, only methods with false compliant (false negative) rates of < 5 % can be used for targeted screening of certain substances in live animals and animal

products [93]. If this requirement was also applied to screening of energetic materials, then appropriate reference materials would be needed to demonstrate this which is not possible for non-target compounds. For suspect and non-target screening methods, perhaps all that can be validated is the ability for the method to detect target or suspect compounds not included in method performance and/or optimisation experiments. That was the approach taken here when assessing method performance and at least 1 ion was detected at 10 ng μL^{-1} for all of the 44 additional suspect compounds considered here.

3.4 Conclusion

Overall, the LC-HRMS method performed well and generalised to a much larger set of target compounds. At least one ion was detected for all 62 compounds studied here. Additionally, multiple oligomers of MEKP were detected, as were mannitol and sorbitol penta- and tetra-nitrates and sorbitol tri- and di- nitrates. For 54 compounds, at least three ions (including the molecular ion/(de)protonated molecule/adduct ion and a fragment ion) could be detected at the highest concentration (10 ng μL^{-1}), even without the use of HCD fragmentation. Although, signal to noise ratios and LODs varied greatly for different compounds and for different ions of the same compound. HRMS analysis was not always able to provide a unique elemental composition, even with a mass accuracy threshold of 1 ppm. Changing the mass accuracy threshold had little effect on the number of elemental compositions for smaller molecules so a mass accuracy threshold of 5 ppm was used here to minimise false negatives. Even if a unique elemental composition could be determined, for all bar two compounds, there were multiple isomers with the same elemental composition. This is a common challenge and it stressed particularly the need for selective separation of isomers with chromatography, or otherwise, to achieve the greatest confidence in identification. In some cases, using multiple ions led to improved selectivity between different explosive compounds and from matrix. However, the greatest improvement in selectivity was seen through the use of a retention time window, highlighting the value of developing a good chromatographic separation. Unlike, the use of multiple ions, utilising the chromatographic separation did not result in poorer LODs. Even with retention time windows there were a few cases of selectivity issues between pairs of isomers on this C_{18} Ar phase, but in the case of 2,4- and 2,6-DNT the use of ion ratios provided sufficient discrimination. The method was also successfully used to screen for explosives, with RDX detected in fingermarks and 12 explosives detected on passive vapour samplers exposed to an explosive magazine. For use as a screening method, the selectivity issues encountered here were not too problematic. However, confirmation by an additional technique or further investigation of potential selectivity issues would be recommended for confirmatory identification. Additionally, full method validation would be required though guidelines are not available yet for suspect and non-target screening LC-HRMS methods.

Chapter 4: Retention time prediction using multiple linear regression and machine learning

4.1 Introduction

Chapter 3 highlighted the importance of chromatographic retention time for the identification of explosives by LC-HRMS. This presents a challenge for the identification of new compounds when reference materials may not be immediately available and in some cases hundreds, or even thousands, of compounds share the same elemental composition. Prediction of retention times using quantitative structure-retention relationships (QSRRs), based on a set of molecular descriptors, may allow for a reduction in the number of compounds for which measured retention times need to be obtained. QSRR models, using different sets of descriptors, have already been used to predict the retention times of small compounds for anti-doping [77,82] and environmental monitoring [80,83,84,90]. One limitation with existing QSRR models is that they are often only applicable to the specific LC conditions used to generate the method. Barron and McEneff demonstrated the generalisability of their artificial neural network (ANN) QSRR model across 10 different instrumental methods and matrices [84]. However, limited diversity in stationary phase was demonstrated, with a C_{18} column used for 9 of the 10 methods and a C_8 column used for the remaining method. Both of these stationary phases can only form Van der Waals forces and so generalisability across reversed phase HPLC columns that can also form π-π or permanent dipole interactions has not yet been shown. Additionally, QSRR models have not yet been used to predict retention times for energetic materials. Therefore, the first aim of this chapter was to investigate the suitability of ANN models using previously selected descriptors for the prediction of explosive retention times for the C_{18} Ar, gradient LC method discussed in Chapter 2. Molecular descriptors were then selected specifically for this LC method and set of analytes and prediction performance was compared. The second aim of this chapter was to use the best prediction models to predict retention times for explosives not included during the training and selection of prediction models. Prediction intervals were calculated to provide a measure of confidence in the predictions and the best two models were used to predict retention times for methyl ethyl ketone peroxide (MEKP) oligomers and mannitol and sorbitol nitrates. Finally, to investigate the potential for the models to reduce the number of compounds for which measured retention times must be acquired, retention times were also predicted for structural isomers with the same elemental composition as three explosives, RDX, TATB and PYX.

4.2 Experimental

4.2.1.1 Acquisition of HPLC retention times

The LC-APCI-HRMS method optimised in Chapter 2 and performance tested in Chapter 3 was used to generate measured retention times for 47 explosives-related compounds. To expand the data set, chromatographic retention times were also obtained for 102 pharmaceuticals using the same column and gradient conditions. A list of the compounds used can be found in the Appendix. For the pharmaceuticals; heated electrospray ionisation, with conditions optimised for pharmaceuticals by Munro et al. [83], was used. A mixed standard, containing all 149 compounds, was analysed (n=6 injections) in both negative and positive mode, with both APCI and ESI to determine average measured retention times which were used to develop a retention time prediction model. Standard deviation was also calculated.

Measured retention times were determined in a fingermark matrix by extracting fingermarks (n=6, 3 male and 3 female donors) in 500 µL of the mixed standard. For measured retention time in pond water, grab samples were used. Following storage at -20 °C in Nalgene bottles, pond water samples were defrosted and acidified with 37% (w/v) hydrochloric acid solution to pH 2. Pond water samples were then filtered under vacuum using Whatman GF/F 0.7 µm glass microfibre filters and the solid phase extraction (SPE) method optimised by Rapp-Wright et al. was used [130]. This involved Oasis HLB (6 mL × 200 mg) cartridges (Waters Corp., Hertfordshire, UK) which were conditioned with 5 mL methanol and washed with 10 mL ultrapure water before 100 mL of the acidified water samples (n=3) were loaded onto the SPE cartridges at 5-10 mL min^{-1}. Cartridges were then washed with 5 mL ultrapure water and dried under vacuum for 10 min before elution with 2.5 mL of mixed standard in acetonitrile. This solution was then injected onto the LC-HRMS.

4.2.1.2 Generation of molecular descriptors

Simplified Molecular Input Line Entry Specification (SMILES) strings were extracted from PubChem [131] and ChemSpider [117], or generated from a molecular structure drawn in ACD/ChemSketch (freeware) [132] if not found in either database. A list of these SMILES strings can be found in the Appendix, Table A. 1. The SMILES were then input into Parameter Client, an open access applet from the Virtual Computational Chemistry Laboratory (VCCLAB) [133] which generated 3209 molecular descriptors using multiple calculation servers including: E-Dragon;

an E-state indices calculation module; two fragment-based indices calculation programs (cfrag and cfrag-I) and ALOGPS which calculated LogP (lipophilicity) and LogS (aqueous solubility). Further information about the number and types of descriptors generated using Parameter Client is summarised in Table 4.1. Percepta PhysChem Profiler (ACD Laboratories, ON, Canada) was used to generate ACD/pKa(acid), ACD/pKa(base), ACD/LogP and ACD/LogD at pH 7 from the SMILES, since pKa and LogD could not be calculated using Parameter Client.

4.2.1.3 Selection of molecular descriptors

Initially whole sets of descriptors with missing values or zero variance for the explosives subset were removed. Where two descriptors had a linear correlation of $R^2 \geq 0.98$, the descriptor least correlated with measured retention time was also removed. Two methods of descriptor selection were used to down select from the remaining 1096 descriptors. The first method involved using the forward selection algorithm in Trajan 6.0 neural network simulator (Trajan Software Ltd., Lincolnshire, UK). This was performed (n=6) for each group of descriptors outlined in Table 4.1. The descriptors from each group, maximum of 10, which gave the smallest network error were then selected and combined with ACD/LogP and ACD/LogD to give 187 descriptors. The second method of selection involved running the genetic algorithm in Trajan 6.0 on all 1096 descriptors (n=10) with a unit penalty of 0.1.

4.2.1.4 ANN modelling

Trajan 6.0 neural network simulator was used for all ANN modelling. The three-step optimisation process outlined by Barron and McEneff [84], was followed for ANN optimisation using descriptors previously selected by Barron & McEneff [84], Goryński et al. [77] and Aalizadeh et al. [80]. As recommended by Barron and McEneff, a 70:15:15 ratio was used to split the dataset into training (n=105), selection (n=22) and test (n=22) sets. During the first phase of optimisation, the datasets were randomly reassigned within these numbers for each network to be tested. The intelligent problem solver tool was run for 15 mins and set to balance network error (based on the selection set) with diversity. At this stage, probabilistic neural networks (PNNs), generalised regression neural networks (GRNNs), radial basis functions (RBFs) and three and four-layer multilayer perceptrons (MLPs) were all evaluated. The best network type was selected from a shortlist of 50 retained networks, based on the lowest consistent prediction errors across the training, selection and test sets. The intelligent problem solver tool was then run for a further 15 min. This time only

Table 4.1: Groups of molecular descriptors generated using Parameter Client [133].

Group of descriptors	Total	Further details
2D autocorrelations	96	Calculated from molecular graph by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag).
3D-MoRSE descriptors	160	Molecular descriptors calculated by summing atom weights viewed by a different angular scattering function.
Atom-centred fragments	120	Molecular descriptors based on the counting of 120 atom-cantered fragments, as defined by Ghose-Crippen. Some fragments are undefined by the authors.
BCUT descriptors	64	Molecular descriptors obtained from the positive and negative eigenvalues of the adjacency matrix, weighting the diagonal elements with atom weights.
Charge descriptors	14	Fourteen charge descriptors, which are reliable only when charges are estimated by quantum molecular method.
Connectivity indices	33	Topological molecular descriptors calculated from the vertex degree of the atoms in the H-depleted molecular graph.
Constitutional descriptors	48	OD-descriptors, independent from molecular connectivity and conformations. Atom and bond counts, molecular weight sum of atomic properties, etc.
Edge adjacency indices	107	Topological molecular descriptors derived from the edge adjacency matrix, which encodes the connectivity between graph edges.
Eigenvalue-based indices	44	Topological descriptors calculated by the eigenvalues of a square (usually symmetric) matrix representing a molecular graph.
E-state Indices	382	Atom-type and bond-type electrotopological state (E-State) indices.
ET-state Properties	3	By-product of the calculation of ET-state Indices. These can be used to check if the molecules were processed correctly by the program.
Functional group counts	121	Molecular descriptors based on the counting of chemical functional groups.
Geometrical descriptors	74	Conformationally dependent descriptors based on the molecular geometry. Reliable values are obtained if reliable conformations were previously calculated.
GETAWAY descriptors	197	Descriptors calculated from the leverage matrix obtained by the centred atomic coordinates (molecular influence matrix, MIM).
GSFRAG	307	Occurrence numbers of special molecular fragments on k=2,...,10 vertices in a molecular graph G
GSFRAG-L	886	Occurrence numbers of special molecular fragments on k=2,...,7 vertices containing one labelled vertex.
Information indices	47	Molecular descriptors calculated as information content of molecules, based on the calculation of equivalence classes from the molecular graph.
Molecular properties	28	Molecular properties calculated from models together with some empirical descriptors.
Randic molecular profiles	41	Derived from the distance distribution moments of the geometry matrix, average row sum of its entries raised at the k^{th} power, normalized by the factor k!
RDF descriptors	150	Molecular descriptors obtained by radial basis functions cantered on different interatomic distances (from 0.5A to 15.5A).
Topological charge indices	21	First 10 eigenvalues (absolute values) obtained from a corrected adjacency matrix.
Topological descriptors	119	Molecular descriptors obtained from molecular graph (usually H-depleted), i.e. 2D-descriptors conformationally independent.
Walk and path counts	47	Molecular descriptors obtained from the molecular graph, counting paths, walks and self-returning walks of different lengths.
WHIM descriptors	99	Statistical indices of the atoms projected onto the 3 principal components obtained from weighted covariance matrices of the atomic coordinates.

the best network type was further evaluated, and the test set fixed according to the best network from the previous stage (but the training and selection sets were randomly resampled). This time the best network architecture was selected from a shortlist of 50 retained networks, again based on the balance of prediction errors across the training, selection and test sets. Finally, the best network type and architecture was replicated (n=6) with the training, selection and test sets fixed.

Optimisation of ANN models using descriptors selected by forward selection or genetic algorithm followed the same general process as above. However, the option for the intelligent problem solver to select a subset of descriptors was used up until the final stage where the best network was replicated. Due to the relatively large number of descriptors at the start of the process, an additional 15 min optimisation stage was used at the beginning and only the descriptors used in the best network (based on the balance of prediction errors across the training, selection and test sets) were taken forward to the next stages of optimisation.

4.2.1.5 Multiple linear regression modelling

Multiple linear regression (MLR) models published by Goryński et al. [77] and Aalizadeh et al. [80] were tested on this dataset by calculating the predicted retention time in Excel using the published equations. Scatter plots for predicted versus measured retention times and prediction error versus retention time were also plotted in Excel. IBM SPSS statistics 24 was used to optimise the MLR models for this LC method based on the training set (n=105 compounds) used for the best ANN model with the same descriptors. Performance of the optimised MLR model on a test set (n=44 compounds) was then tested in Excel as described above.

4.2.1.6 Statistical analysis

Excel was used to calculate R^2 between descriptors, plot all scatter plots and calculate R^2, slope and gradient for measured versus predicted retention times. Mean absolute error (MAE) and root mean squared error (RMSE) were also calculated in Excel, along with percentiles, interquartile range and total range for prediction errors. SPSS was used to assess the statistical significance of Pearson's correlation and Spearman's rank order correlation. SPSS was also used for a paired T-test, between the residual errors of the best models, the Shapiro-Wilk test for normality and generation of normal Q-Q plots. Stacked histograms for the training, selection and test errors were plotted using Excel. Performance intervals were also calculated in Excel, the details of which are discussed further in Section 4.3.4.3.

4.3 Results and discussion

4.3.1 Building a dataset of measured retention times

Average measured retention times (n=6) were initially generated for 47 compounds relevant to the forensic analysis of explosives using the LC-APCI-HRMS method developed in Chapter 2 and performance tested in Chapter 3. In addition to predicting the retention times of explosives, the ability to predict the retention times of non-explosives with the same elemental composition was desirable for exclusion purposes and to expand the chemical diversity of the model. Therefore, retention times were also obtained for 102 drugs and pharmaceuticals using the same LC method but with ESI conditions optimised for drugs by Munro, during a previous study [83]. This gave a combined dataset of 149 compounds which was still a relatively small dataset in comparison to the 550 compounds used by Bade et al. [90]. Barron & McEneff investigated the minimum number of cases required for ANN learning and found that, whilst prediction accuracy improved when larger datasets were used for training, a network trained using 36 compounds was still able to provide good prediction accuracy (average absolute error in retention time, $\Delta t_R = 1.15 \pm 1.05$) for the blind test set [84].

4.3.2 Developing prediction models with molecular descriptors from previous studies

Three sets of molecular descriptors from published QSRR studies (Table 4.2), were investigated here for prediction of retention times for the gradient LC separation developed using a C_{18} Ar column in Chapter 2. These included the descriptors used by: Goryński et al. for separation of illicit drugs, metabolites and other anti-doping compounds on a PFP column [77]; Barron & McEneff for separation of pharmaceuticals, drugs of abuse, biocides and industrial chemicals, using C_{18} and C_8 columns [84] and Aalizadeh et al. for separation of pesticides and other emerging contaminants detected in negative ionisation mode on a C_{18} column [80]. The only descriptor used in all three models was the Ghose–Crippen logarithmic octanol–water partition coefficient (ALOGP). Although, the logarithmic octanol water distribution coefficient (LOGD) was also used by both Barron & McEneff and Aalizadeh et al. Due to the pH dependant nature of LogD, here it was calculated at pH 7 to reflect the mobile phase used for this LC-HRMS method, despite Barron & McEneff using LogD at pH 5.2 and Aalizadeh et al using LogD at pH 6.2 for their models.

Table 4.2: Molecular descriptors used in previous retention time prediction models

Descriptor	Definition	Goryński et al. 2013	Barron & McEneff 2016	Aalizadeh et al. 2016
ALOGP	Ghose–Crippen octanol–water partition coefficient	X	X	X
MLOGP	Moriguchi octanol–water partition coefficient	-	X	-
LOGD	Log of the octanol water distribution coefficient	-	X	X
BLTA96	Verhaar baseline toxicities for Algae (mmol L−1)	-	-	X
Ui	Unsaturation index	-	X	-
Hy	Hydrophilic factor	-	X	-
R2p	R autocorrelation of lag 2, weighted by atomic polarizabilities,	X	-	-
HATS6m	Leverage-weighted autocorrelation of lag 6/weighted by mass	-	-	X
BELe6	Lowest Burden eigenvalue 6, weighted by atomic Sanderson electronegativities	X	-	-
BEHm4	Highest Burden eigenvalue 4, weighted by atomic masses	-	-	X
EEig14r	Eigenvalue 14 from edge adj. matrix weighted by resonance integrals	-	-	X
CIC1	Complementary Information Content index (neighbourhood symmetry 1-order)	-	-	X
nDB	Number of double bonds	-	X	-
nTB	Number of triple bonds	-	X	-
nC	Number of carbons	-	X	-
nO	Number of oxygens	-	X	-
nR04	Number of 4 membered rings	-	X	-
nR05	Number of 5 membered rings	-	X	-
nR06	Number of 6 membered rings	-	X	-
nR07	Number of 7 membered rings	-	X	-
nR08	Number of 8 membered rings	-	X	-
nR09	Number of 9 membered rings	-	X	-
nBnz	Number of benzene-like rings	-	X	-

X = Descriptor included in model

4.3.2.1 Artificial neural networks with molecular descriptors from previous studies

In order to evaluate the suitability of the three sets of molecular descriptors for predicting the gradient retention time of explosives on a C_{18} Ar stationary phase, neural networks were optimised for each set of descriptors using Trajan 6.0 and the three-step optimisation process outlined by Barron & McEneff [84]. The correlation between predicted and measured retention times for the best neural networks using each set of descriptors are shown in Figure 4.1 A, with the prediction error plotted against retention time in Figure 4.1 B. Three-layer multilayer perceptrons (MLPs) provided the best neural network models for the 16 descriptors used by Barron & McEneff and the 7 descriptors used by Aalizadeh, whereas a radial based function (RBF) provided the best model for the 3 descriptors used by Goryński et al. Statistically significant negative correlation ($p<0.001$) was seen between retention time and prediction error for all three models, with over prediction of more early eluting compounds and under prediction of more later eluting compounds,

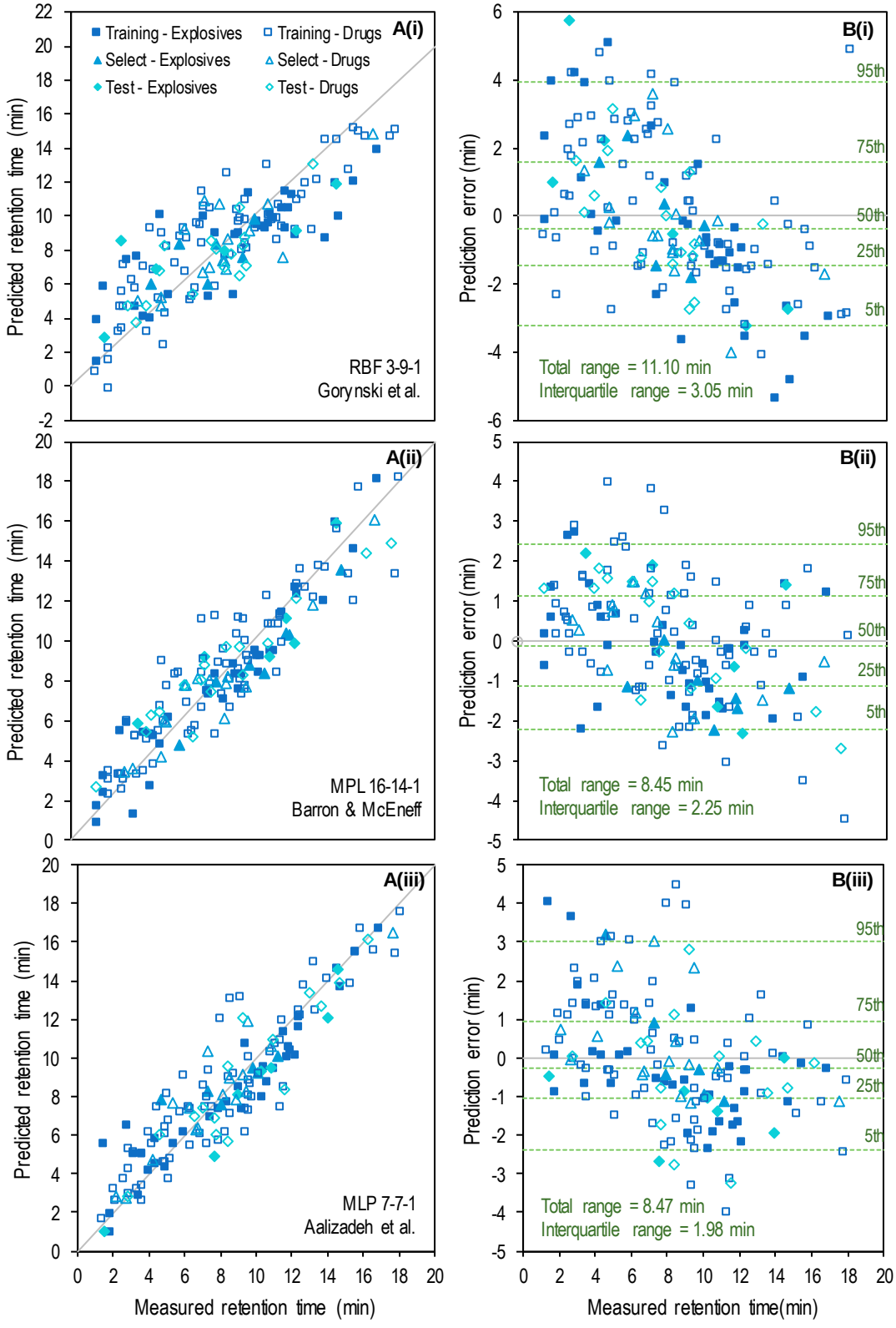


Figure 4.1: Correlation between measured and predicted retention times (A) and prediction error (B) for the best models using three different sets of molecular descriptors. Grey lines indicate perfect prediction and dashed green lines indicate percentiles for the complete dataset.

indicating that the models were not fully explaining the retention mechanism. This trend was strongest for the model using the Goryński et al. descriptors with a Pearson correlation coefficient (R) of -0.559, compared to -0.437 and -0.399 for the models using Barron & McEneff and Aalizadeh descriptors respectively.

When optimising each neural network, the data was randomly divided into three data sets: training set (n=105) which as the name suggests was used to train the model based on the known retention times; selection set (n=22) which was used during the optimisation process to control for over fitting and the test set (n=22) which was not involved in network optimisation but was also considered when selecting the best network. Due to the random assignment of compounds into training, selection and test sets during the first stage of ANN optimisation, the test set was not consistent for the models using different molecular descriptors. Additionally, that test set did not always have the largest errors. This was perhaps unsurprising given that this was an internal test set and consistency in errors across training, selection and test sets was a network selection criterion. Therefore, it was important to consider the errors across all datasets and not just the test set error when evaluating performance of the models. For all models the training, selection and test sets each contained a mixture of explosives and drugs. To ensure that the models were performing well for both explosive and non-explosive compounds the results were also analysed for explosive and drug subsets individually as well as for the combined dataset.

Performance statistics, including: R^2, slope and intercept of lines of best fit are shown in Table 4.3. Across all subsets the model using the descriptors from Goryński et al. showed the poorest correlation between measured and predicted retention times with the lowest R^2 values and the slopes and intercepts furthest from 1 and 0, which would be the case if x=y and the predicted retention time was equal to the measured retention time. The model using the descriptors from Barron & McEneff had the highest R^2 values for all subsets apart from the explosives in the test set. Whereas the model using the descriptors from Aalizadeh et al. had the slope and intercept closest to 1 and 0 for the test sets, but the slope and intercept furthest from 1 and 0 for the explosives in the selection set.

Table 4.3 also shows the root mean square error (RMSE) and mean absolute error (MAE) for the models using each set of descriptors. The RMSE for the combined test set was very similar for models using the descriptors selected by Barron & McEneff and Aalizadeh et al. (1.52 and

Table 4.3: Performance of the best ANN for each set of descriptors

Subset	Goryński et al. (RBF:3-9-1)					Baron & McEneff (MLP:16-14-1)					Aalizadeh et al. (MLP:7-7-1)				
	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)
Training - Explosives	0.67	0.55	3.57	2.53	2.00	0.91	0.91	0.54	1.28	1.05	0.90	0.82	1.24	1.48	1.12
Training - Drugs	0.75	0.79	2.05	2.21	1.86	0.83	0.83	1.53	1.63	1.27	0.81	0.84	1.33	1.74	1.38
Training - Combined	0.71	0.71	2.53	2.32	1.91	0.86	0.86	1.17	1.52	1.20	0.84	0.83	1.32	1.66	1.29
Selection - Explosives	0.39	0.44	4.39	1.51	1.32	0.94	0.91	-0.23	1.38	1.23	0.71	0.36	5.67	1.49	1.16
Selection - Drugs	0.61	0.68	2.66	1.86	1.39	0.89	0.83	1.22	1.16	1.00	0.88	0.88	1.32	1.28	0.95
Selection - Combined	0.59	0.64	2.89	1.77	1.37	0.90	0.81	1.18	1.23	1.07	0.84	0.81	1.83	1.35	1.01
Test - Explosives	0.65	0.46	4.46	3.08	2.57	0.76	0.73	2.86	1.78	1.69	0.96	1.00	-1.16	1.46	1.20
Test - Drugs	0.65	0.68	2.37	1.58	1.30	0.93	0.73	2.46	1.40	1.26	0.84	0.91	0.59	1.53	1.13
Test - Combined	0.62	0.56	3.40	2.10	1.65	0.89	0.74	2.50	1.52	1.38	0.87	0.94	0.02	1.51	1.15
All	0.69	0.69	2.62	2.22	1.79	0.86	0.84	1.35	1.48	1.21	0.84	0.84	1.26	1.59	1.23

RMSE, root mean square error; MAE, mean absolute error.

1.51 min, respectively), whereas the model using the descriptors selected by Goryński et al. had a RMSE of 2.10 min. Interestingly this was still lower than the RMSE of 3.45 min reported by Goryński et al. for predicting the retention times on a PFP column using their multiple linear regression model [77]. In contrast, Aalizadeh et al. reported a much lower RMSE of 0.40 min for their best nonlinear model for predicting retention times on a C_{18} column [80]. The MAE of the combined test set, which is less affected by outliers than the RMSE, was also largest for the model using the descriptors from Goryński et al. (MAE = 1.65 min), followed by the model using the descriptors from Barron & McEneff with MAE of 1.38 min. This was greater than the best performing model reported by Barron & McEneff for prediction of retention times on a C_{18} column, which had an MAE of 0.37 min, but lower than the poorest performing model which had an MAE of 1.82 min [84]. The fact that the model using only three descriptors, selected by Goryński et al., had the largest prediction errors suggested that those three descriptors led to an over simplified prediction model. Similarly, the larger ANN prediction errors, using descriptors selected by Barron & McEneff, for the C_{18}Ar column used here than for the C_{18} and C_8 columns used in the original paper, suggested that the descriptors may not fully describe retention on a C_{18} Ar column. In addition to the van Der Waals forces formed on a C_{18} or C_8 column, π-π interactions are also possible on a C_{18} Ar column.

4.3.2.2 Linear prediction models with molecular descriptors from previous studies

While Barron & McEneff used non-linear artificial neural networks (ANN) [84], Aalizadeh et al. used both linear (multiple linear regression, MLR) and non-linear (ANN and support vector machines, SVM) models [80] and Goryński et al. used a linear MLR model [77]. Therefore, linear MLR models were also investigated for the prediction of gradient retention times on a C_{18} Ar column. Initially the MLR models developed by Goryński et al. (t_R = 9.67R2p + 3.67BELe6 + 1.70ALogP) [77] and Aalizadeh et al. (t_R = - 0.4297+ 0.6242logD + 0.4649AlogP - 0.08647BLTA96 - 0.6998EEig14r + 0.7589CIC1 + 1.551BEHm4 + 0.7907HATS6m) [80] for their respective LC methods were tested on this dataset using the C_{18} Ar column to check for generalizability.

The scatter plots for predicted versus measured retention times (Figure 4.2A(i)) and prediction error over time (Figure 4.2B(i)) showed some separation between drugs and explosives for the Goryński et al. MLR model, with over prediction of all drug retention times but under prediction of most explosive retention times. The Aalizadeh et al. MLR model on the other hand showed strong

negative correlation (R^2 = 0.81) between the prediction error and retention time for both the drugs and explosives subsets (Figure 4.2B(ii)). For both MLR models the RMSE and MAE (Table 4.4) were greater than for the best ANN models using the same molecular descriptors (Table 4.3). Additionally, the RMSE were greater for this dataset than for the test sets presented in the original publications, with a combined RMSE of 5.32 min, compared to 3.45 min [77], for the Goryński et al. model and a combined RMSE of 2.75 min, compared to 1.22 min [80], for the Aalizadeh et al. MLR model.

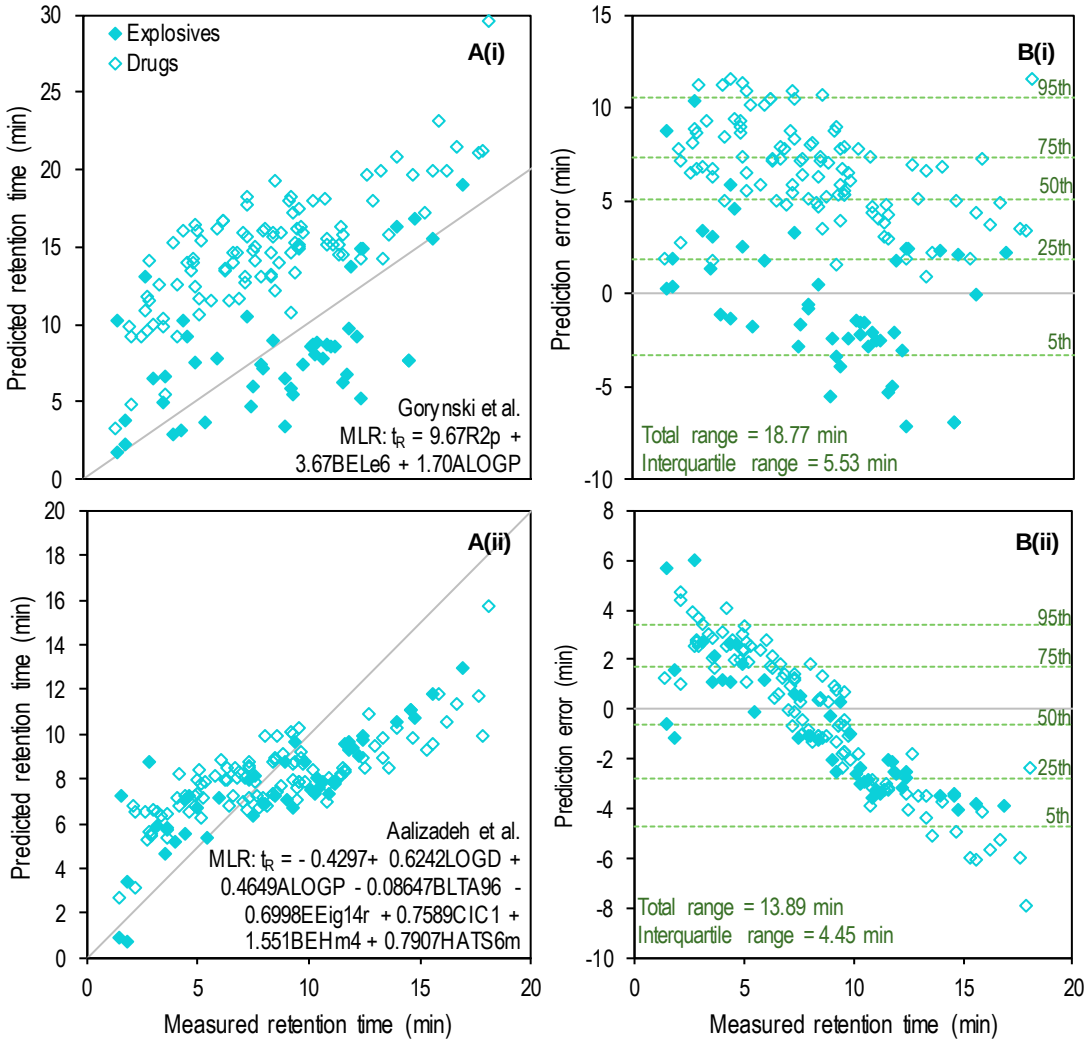


Figure 4.2: Performance of previous MLR models on this dataset with A) showing the correlation between predicted and measured retention times and B) showing the prediction error across the run. Grey lines indicate perfect prediction and dashed green lines show percentiles.

Table 4.4: Performance of published MLR models on this dataset.

	Goryński et al. MLR: t_R = 9.67R2p + 3.67BELe6 + 1.70ALOGP					Aalizadeh et al. MLR: t_R = - 0.4297+ 0.6242LOGD + 0.4649ALOGP - 0.08647BLTA96 - 0.6998EEig14r + 0.7589CIC1 + 1.551BEHm4 + 0.7907HATS6m				
Subset	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)
Explosives	0.32	0.53	3.50	3.78	2.99	0.71	0.49	3.41	2.63	2.28
Drugs	0.60	0.74	8.71	5.22	6.52	0.62	0.34	5.17	2.80	2.33
Combined	0.27	0.63	7.37	5.32	5.40	0.63	0.39	4.62	2.75	2.32

The two previous MLR models did not generalise well to this LC method. However, given that retention times, and therefore the coefficients included in an MLR model, are affected by the LC conditions used (e.g. flow rate, mobile phase composition, stationary phase, particle size, column length and diameter); it was unsurprising that the MLR models performed worse when tested with a dataset using a different LC method and stationary phase. Neither was it surprising that MLR models trained using retention times from a different LC-method performed worse than the neural networks which were trained using measured retention times for this LC method. For a fairer comparison, multiple regression analysis was performed in SPSS, for both the Goryński et al. and Aalizadeh et al. descriptors. MLR models were trained using the same training sets (n=105) as the respective ANN model, before the remaining compounds (n=44) from the select and test sets were used to assess prediction. Performance of the optimised MLR models are shown in Figure 4.3 and Table 4.5.

The MLR model trained using the Goryński et al. descriptors had the following equation:

$$t_R = 9.334(\pm1.020) - 11.546(\pm2.047)R2p + 5.683(\pm0.203)BELe6 + 1.954(\pm0.203)ALOGP.$$

Multiple regression analysis showed that the three variables (R2p, BELe6 and ALOGP) predicted the retention time with statistical significance ($p < 0.0005$, $R^2 = 0.615$) and each of the three variables added significantly to the prediction ($p < 0.0005$). Figure 4.3 shows that including a mixture of explosives and drugs in the training set removed the separation between the two subsets witnessed in Figure 4.2. Optimisation of the coefficients for this dataset improved predictive accuracy, with the median error (50th percentile) reduced to -0.11 min in Figure 4.3 B(i), compared to 5.07 min in Figure 4.2 B(i) and the mean bias error (MBE) reduced from 4.30 to 0.27 min. The maximum and average errors were also reduced, indicating improved precision, with the combined test set RMSE of 2.95 min (Table 4.5), lower than the RMSE of 3.45 min published

Table 4.5: Performance of MLR models optimised for this dataset.

Subset	Goryński et al. MLR: t_R = 9.334 -11.546R2p + 5.683BELe6 + 1.954ALOGP					Aalizadeh et al. MLR: t_R = -1.933 + 0.131LOGD - 0.747ALOGP – 1.560 BLTA96 +0.507EEig14r +0.822CIC1 +1.464BEHm4 +2.422HATS6m				
	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)
Training - Explosives	0.68	0.54	3.30	2.61	2.24	0.60	0.54	3.58	2.76	2.38
Training - Drugs	0.63	0.66	3.43	2.63	2.21	0.09	0.11	7.36	3.86	3.02
Training - Combined	0.62	0.62	3.40	2.62	2.22	0.27	0.27	6.06	3.54	2.81
Test - Explosives	0.33	0.36	4.90	2.99	2.46	0.32	0.45	3.95	3.19	2.36
Test - Drugs	0.24	0.43	5.29	2.94	2.46	0.26	0.22	7.10	3.22	2.57
Test - Combined	0.25	0.40	5.28	2.95	2.46	0.23	0.27	6.36	3.17	2.46
All	0.54	0.57	3.86	2.72	2.29	0.26	0.27	6.13	3.44	2.71

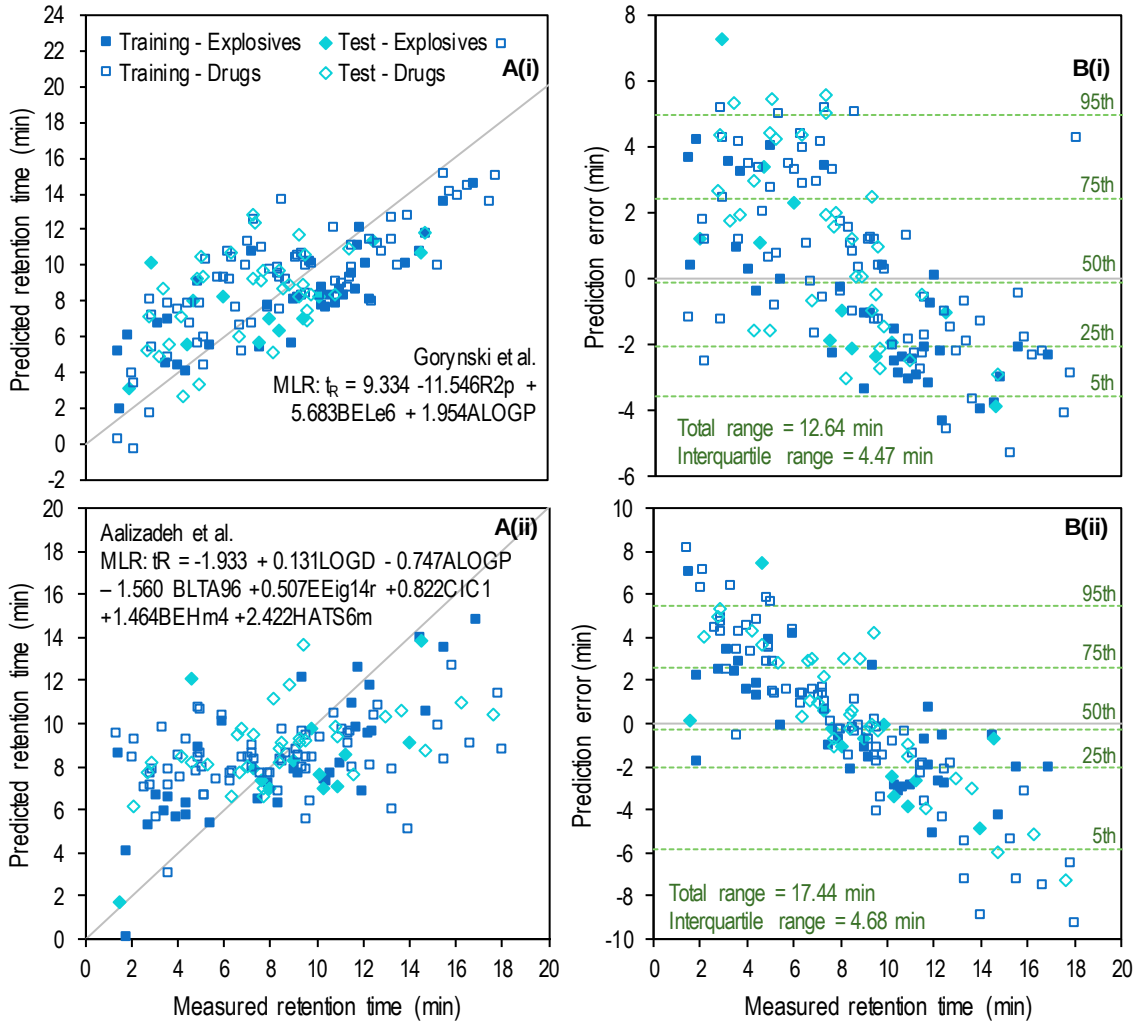


Figure 4.3: Performance of MLR models developed for this dataset with A) showing correlation between predicted and measured retention times and B) showing the prediction error over time. Grey lines indicate perfect prediction.

in the original paper, but not as low as the combined test set RMSE of 2.10 min for the ANN model using these descriptors (Table 4.3).

The MLR model trained using the Aalizadeh et al. descriptors had the following equation:

$$t_R = -1.933(\pm2.812) + 0.131(\pm0.305)LOGD - 0.747(\pm0.535)ALOGP - 1.560(\pm0.644)BLTA96 + 0.507(\pm0.714)EEig14r + 0.822(\pm0.762)CIC1 + 1.464(\pm0.895)BEHm4 + 2.422(\pm1.375)HATS6m.$$

Multiple regression analysis showed that these seven variables (LOGD, ALOGP, BLTA96, EEig14r, CIC1, BEHm4 and HATS6m) predicted the retention time with statistical significance ($p < 0.0005$). However, BLTA96 was the only descriptor which had a significant contribution to the model ($p<0.05$). As shown in Figure 4.3A(ii), correlation between predicted and measured retention times was also low. The coefficient of determination (R^2) was 0.23 for combined test set, indicating that only 23% of the variance in retention time was explained by this model. Optimising the coefficients did reduce prediction bias with the median errors reduced from -0.62 to -0.25 min and the MBE reduced from -0.49 to 0.002 min. However, in this case the absolute prediction errors for the combined test set were even larger than for the MLR model optimised for Aalizadeh et al.'s LC method (RMSE = 3.17 min versus RMSE = 2.75 min, MAE = 2.46 min versus MAE = 2.32 min). In both cases, the best ANN model had lower errors (RMSE = 1.51 min, MAE = 1.15 min).

Overall, non-linear, ANN models performed better than linear MLR models for the descriptors investigated here. ANN models using the descriptors selected by Barron & McEneff and Aalizadeh et al. performed better than the ANN model using descriptors selected by Goryński et al. However, in both cases performance was not as good as the best model presented in the original papers. This may be due to differences between the C_{18} Ar stationary phase used here and the C_{18} or C_8 phases used in the previous studies. Descriptors selected to predict retention on a C_{18} or C_8 column, where Van der Waals forces are the only interactions, may not fully describe retention on a C_{18} Ar column where there is the potential for additional interactions (e.g. π-π interactions) with the stationary phase.

4.3.3 Selecting molecular descriptors for retention time prediction on a C_{18} Ar column

Using a free online software (VCCLAB, Parameter Client [133]), 3224 molecular descriptors were easily generated using simplified molecular input line entry systems (SMILES) to describe molecular structure. Additionally, ACD/LogP, pKa and pH-dependant ACD/LogD(pH=7) were

generated using ADC Labs PhysChem profiler to give a total of 3227 descriptors. Too many independent variables, in this case molecular descriptors, can cause problems for prediction models. For example, if there are enough molecular descriptors to describe each molecule in the training set uniquely then the model will suffer from overtraining and poor generalisability, since the model would no longer be working from trends. Larger numbers of descriptors also require increased computational time due to increased complexity [134]. Therefore, a selection process was required to reduce the number of molecular descriptors.

4.3.3.1 Reducing the number of descriptors

Initially, descriptors containing very little or no information were removed. Removal of descriptors with missing values reduced the number of descriptors from 3224 to 2788. Then descriptors with zero variance across the data set were removed, reducing the number of descriptors to 1900; followed by all descriptors with zero variance for the explosives subset which further reduced the number of molecular descriptors to 1515. Four of the descriptors used by Barron & McEneff (nTB, nR04, nR05 and nR07) were eliminated at this stage due to having zero variance for the explosives subset.

Highly correlated descriptors encode the same underlying information [134] resulting in limited value from including multiple highly correlated descriptors in a model. While some neural networks can ignore useless or redundant variables (e.g. multilayer peceptrons), others (e.g. radial based functions) cannot and are adversely affected [86]. Therefore, where two descriptors were highly linearly correlated (defined here as R^2>0.98), the descriptor least correlated with retention time was removed. This reduced the number of molecular descriptors to 1096. Two more of the descriptors used in previous methods, MLOGP and BLTA96 were removed at this stage due to high linear correlation (R^2> 0.9999) with the Verhaar model of Daphnia base-line toxicity from MLOGP (BLTD48), as shown in Figure 4.4. A third descriptor based on the Verhaar model of base-line toxicity from MLOGP, but this time in Fish (BLTF96), was also highly correlated with BLTD48 (again, R^2> 0.9999) and so also removed.

4.3.3.2 Linearity between descriptors and measured retention time

Before developing a neural network, linearity between the measured retention times and each molecular descriptor was examined. Initially this was performed by calculating R^2 values for the 1096 descriptors remaining, after descriptors with missing values, zero variance or high

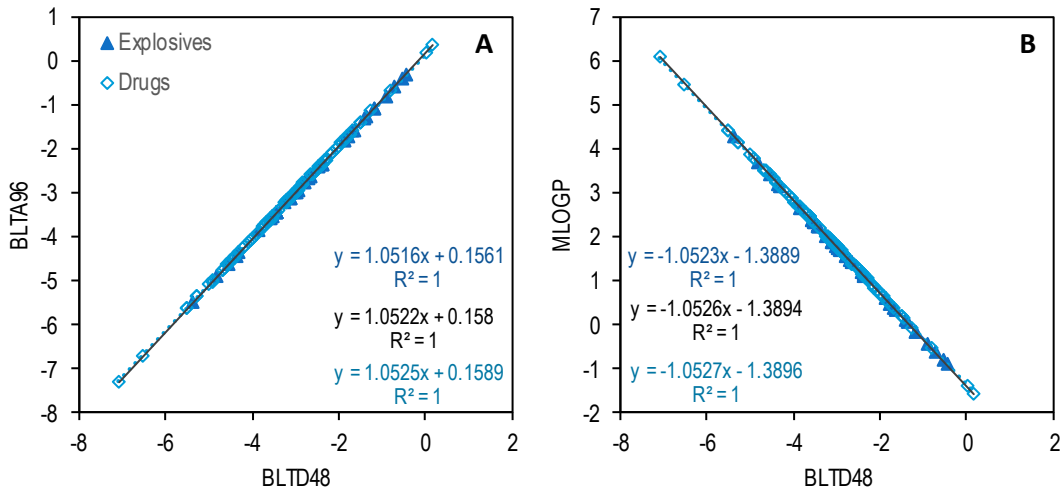


Figure 4.4: Collinearity between A) BLTD48 and BLTA96 and B) BLTD48 and MLOGP. Lines of best fit are shown in dark blue for explosives, light blue for drugs and black for the combined dataset.

correlation had been removed. Scatter plots were then examined for the descriptors with the greatest R^2 values (Figure 4.5). The three descriptors showing greatest linearity with measured retention time for the explosives subset were ACD/LogD at pH 7, ACD/LogP and BLTD48 (Figure 4.5 A-C). ACD/LogD at pH 7 and ACD/LogP, two descriptors calculated by ACD Labs Percepta software, were also within the top five descriptors for the drugs subset. ACD/LogD and ACD/LogP showed positive correlation with retention time, as expected for reversed-phase chromatography where more hydrophobic compounds are retained longer on the non-polar stationary phase. BLTD48 showed negative correlation with retention time. This is because toxicity is related to lipophilicity, due to bioavailability, with low values of BLTD48 indicating high toxicity and therefore high lipophilicity which results in longer retention times by reverse-phase chromatography.

ACD/LogP is a fragment-based approach for predicting the partition coefficient (LogP) from a compounds chemical structure [135]. LogP describes the partition of the unionised form of a compound between octanol and water at equilibrium. Two other methods for predicting LogP were also used to generate the initial set of molecular descriptors: Ghose–Crippen LogP (ALOGP) which is an atom based approach [136] and Moriguchi LogP (MLOGP) which is based on topological descriptors [137]. In most cases the three different prediction methods resulted in different LogP values which leaves the question of which was the most accurate LogP prediction? Manhold et al. found performance of ACD/LogP > ALOGP > MLOGP for predicting Log P using a public dataset but ACD/LogP did not perform as well as ALOGP or MLOGP for their in-house datasets [138]. For this study, accuracy of LogP prediction was less of a priority, but it was

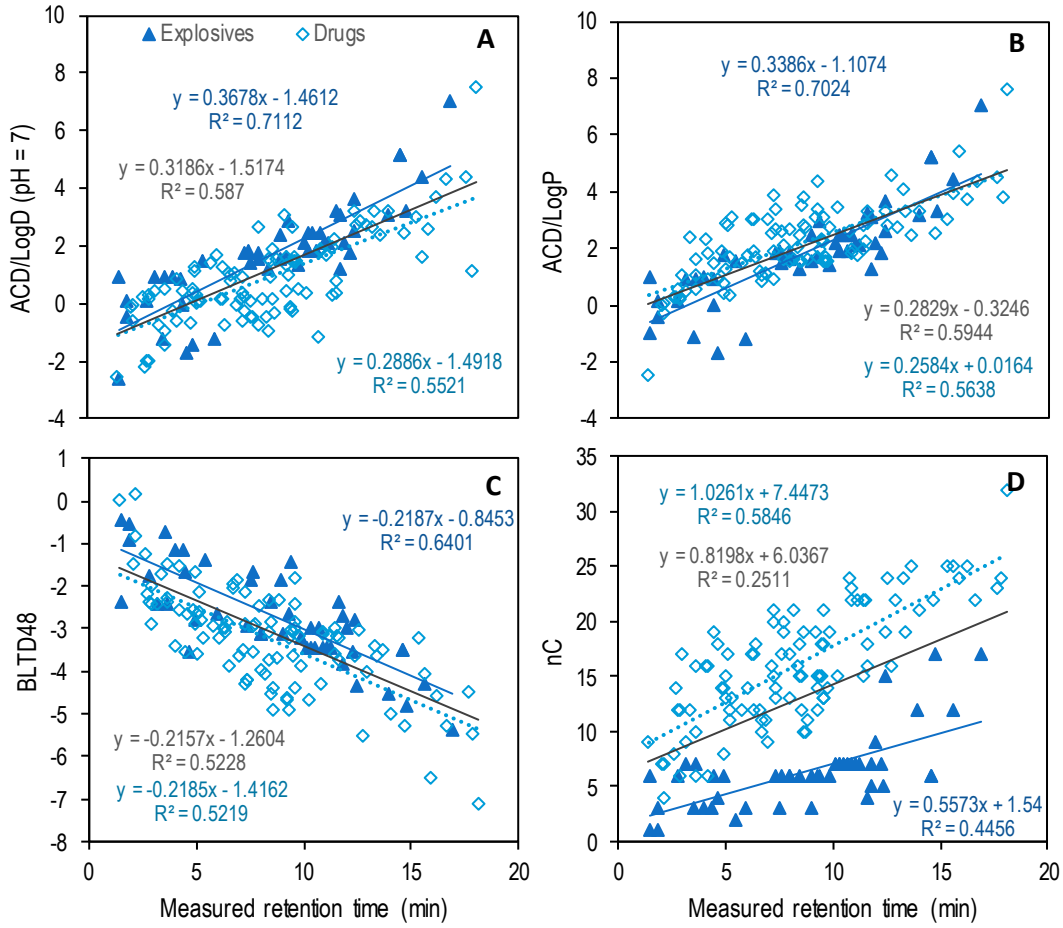


Figure 4.5: Linear correlation between measured retention time and A) ACD/LogD at pH 7, B) ACD/LogP C) BLTD48 and D) number of carbon atoms (nC). Lines of best fit are shown in dark blue for explosives, light blue for drugs and black for the combined dataset.

important to remember that these were only predicted LogP values when interpreting the correlation with retention time. LogP also only considers unionised compounds, so for ionised compounds the pH dependent distribution coefficient (LogD), which describes the distribution of ionised and unionised compounds between octanol and water, may be more appropriate.

The majority of explosives were neither acidic nor basic and so ACD/LogD was equal to ACD/LogP as shown in Figure 4.6 A. The two main exceptions to this were picric acid (PA), with a pK_a (acid) of 0.62 (LogD (pH=7) = LogP - 3.15), and nitroguanidine (NQ), with a pK_a (acid) of 5.23 (LogD (pH=7) = LogP - 1.58). HMTD, with pK_a (base) of 5.28, also had a slight difference between ACD/LogD and ACD/LogP (–0.01) but very little would have been ionised at the relevant pH for this LC method. For the drugs subset (Figure 4.6 B), 72 out of 104 compounds had a LogD <LogP at pH 7, indicating ionisation. This was perhaps unsurprising given that drugs often contain acidic and/or basic groups. What was perhaps more surprising was that ACD/LogP had a slightly

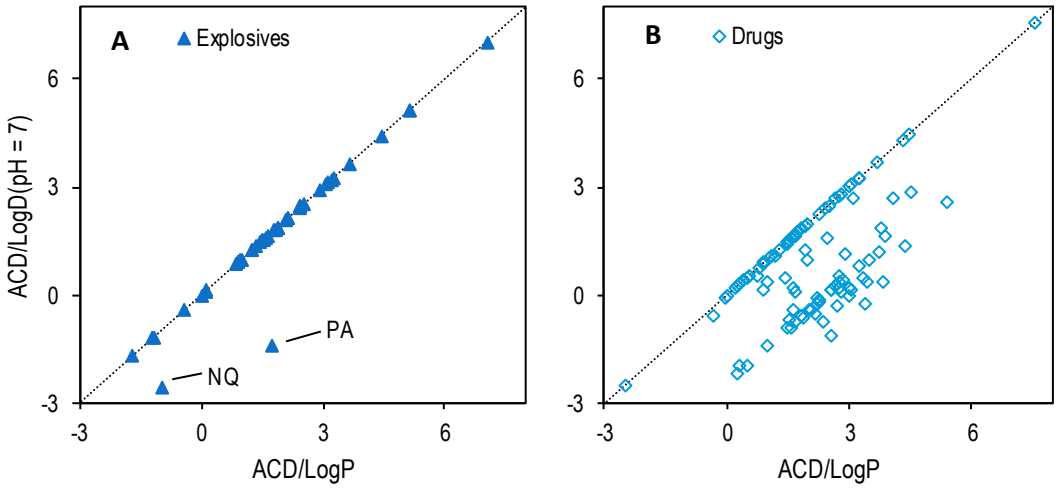


Figure 4.6: Correlation between LogP and LogD at pH 7 for A) explosives and B) drugs with the dashed line showing x=y.

greater correlation with retention than ACD/LogD for the drugs subset and the combined dataset, but ACD/LogD had greater correlation with retention time for explosives (Figure 4.5). Use of LogD might be more appropriate than LogP, provided accurate pH and pK_a values are known, due to the consideration of ionisation which will affect retention. For the majority of explosives consideration of ionised compounds was not important and so LogP might be a suitable alternative if only interested in the prediction of retention times for organic, high-order explosives. However, consideration of ionised compounds will likely be important for exclusion purposes when generalisability to non-explosive compounds is required.

The descriptor with the greatest linear correlation with retention time for the drugs subset was the number of carbon atoms (nC), shown in Figure 4.5 D. The number of carbons showed poor linearity (R^2 = 0.2511) for the combined data set due to some separation between the explosives and drugs subsets. While the majority of drugs had more than 10 carbon atoms, the majority of explosives contained less than 10 carbon atoms, with 15 explosives containing 6 carbon atoms and 12 explosives with 7 carbon atoms. The explosives subset also showed greater chromatographic retention of compounds with fewer carbon atoms. This was likely due to a combination of the drugs dataset containing more acidic/basic groups which would favour interactions with the polar mobile phase and the explosives dataset containing lots of aromatic compounds which would form π-π interactions with the C_{18} Ar stationary phase. This difference

between the two subsets highlights the importance of including explosives in the training set when developing a model for the prediction of explosive retention times, rather than simply using a model developed for drugs.

Of course, linear correlation is not the only form of correlation and it is possible that other descriptors would fit better to polynomial or logarithmic equations. Additionally, with a non-linear prediction model, descriptors may be transformed and/or combined when incorporated into the model. This could lead to interdependencies, where variables may be of limited use individually but valuable in combination with another variable. Therefore, poor linear correlation with retention time does not inherently mean a descriptor would not be of value for modelling retention times and so descriptors were not eliminated due to poor linearity.

4.3.3.3 Selection of independent variables

Once descriptors with missing values, zero variance and high collinearity ($R^2 > 0.98$) were removed, 1096 descriptors remained; too many for use in a prediction model and so further selection was required. In their review paper, Yousefinejad and Hemmateenejad highlighted the importance of variable selection and discussed a number of classical and artificial intelligence methods for selection [87]. The neural network software used here, Trajan 6.0, provided the option of three different variable selection algorithms: forward selection, backward selection and genetic algorithm [86]. With forward selection, independent variables are introduced one at a time, starting with the variable that gives the best prediction on its own, followed by the variable that most improves the prediction, until the stopping criteria is reached [87]. Starting with the variables that provide the best prediction is an efficient method for selecting from a large number of variables, as is the case here. However, once a variable enters the model, it cannot be removed. This means combinations without previously selected descriptors are not tested, which is one of the main disadvantages of using forward selection [87]. With backward selection, initially all independent variables are included in the model and they are removed one by one according to selection criteria until either all remaining variables are significant, or only one variable remains. Backwards selection suffers from a similar disadvantage to forward selection in that not all combinations are tested [87]. However, backwards selection does keep important interdependent variables, that will not be included by forward selection unless they also improve the model individually. The process is more time consuming than forwards selection though, especially when there are large

numbers of variables as is the case here, since all variables are included in the initial evaluations for backwards selection [86]. Large numbers of variables also make it difficult for the backwards selection algorithm to work well, especially if there are only a few weakly predictive variables in the set [86]. Additionally, backwards selection tends to select a higher number of variables than required which can lead to over-fitted models [87].

The genetic algorithm is an artificial intelligence method of variable selection based on natural selection in biological evolution [87]. Independent variables are represented as binary strings with a 0 indicating that a variable should not be used and a 1 indicating that it should be used [86]. The genetic algorithm randomly generates an initial population of strings and a portion of this population is selected, using a fitness function, for transfer to each successive generation, resulting in improved strings. Cross-overs (where two strings swap ends) and mutations (where some variables are randomly flipped from 0 to 1 or vice versa) vary the combination of variables future generations inherit from their parents. The algorithm terminates once the maximum number of generations, time or a plateau in the fitness function is reached [87]. The main advantage of using genetic algorithms for feature selection is their ability to recognise interdependent variables, especially when located close to each other on the string [86]. Genetic algorithms are relatively time consuming but essentially unaffected by the number of variables [86], making them a good option for selecting from the large number of molecular descriptors available here.

Due to the large number of independent variables (1096 molecular descriptors) from which to select, and the disadvantages of using backwards selection for large numbers of molecular descriptors only forwards selection and genetic algorithm feature selection methods were used here. After the use of feature selection methods, the number of descriptors was reduced even further by allowing a subset of variables to be selected during the first stage of neural network training and optimisation. This meant that the final selection of descriptors was performed using the type of neural network used for the actual prediction model, rather than the generalised regression neural network (GRNN) used by the feature selection algorithms due to the speed of training [86].

4.3.3.3.1 Forwards feature selection for ANN retention time prediction

Attempts to run the forward selection algorithm on all 1096 descriptors at the same time repeatedly resulted in an error as the software was unable to cope with so many descriptors.

Therefore, the forward selection algorithm was run on each group of descriptors individually and the 10 best descriptors from each group (based on smallest network error) were selected. Each time the forward selection algorithm was run (n=6 for each group of descriptors), differences were seen in the descriptors selected. Given that in theory when using the forward selection algorithm descriptors should be added to the model in order of the impact they have on prediction, this was perhaps surprising and may indicate the presence of interrelated descriptors. It also highlights the randomness of the feature selection process. Where more than 10 descriptors were selected for each group, only the 10 descriptors leading to the lowest network error were taken onto the next stage to keep the number of descriptors low. Even so, 187 descriptors were taken through to the next stage, use of the Intelligent Problem Solver function to select subsets of these descriptors for optimisation of neural networks. A subset of 32 descriptors (Table 4.6) gave the best prediction model, a 3-layer multilayer perceptron (MLP:32-2-1). Sensitivity analysis was performed to test how well the model performed without each descriptor in turn and the ratio of network errors without/with the descriptor was calculated and used to order the descriptors shown in Table 4.6. Removal of LogD had the biggest effect, increasing the network error by 1.312 times. At the other end of the table, p1p1p2-6N had an error ratio of <1 indicating that the network error was reduced by its removal. However, attempts to retrain a model without this descriptor did not result in improved performance. Sensitivity analysis only evaluates the importance of each descriptor within the context of a particular network and it is not possible to fully link sensitivity analysis with the importance of each descriptor for predicting retention time more generally [86]. Here, an average of 6 networks with fixed datasets and the same architecture, but different optimised weightings associated with each, was used to account for some of the variability in sensitivity analysis between networks. However, interdependent and correlated descriptors would still affect the results of sensitivity analysis.

Table 4.6: Subset of 32 forward selected descriptors and sensitivity of MLP:32-2-1 to these descriptors

Descriptor	Definition	Group	Sensitivity*	SD**
LogD (pH = 7)	Log of the octanol water distribution coefficient at pH 7 predicted by ACD Labs.	Molecular properties	1.312	0.135
GATS1m	Geary autocorrelation - lag 1 / weighted by atomic masses	2D autocorrelations	1.149	0.070
BLTD48	Verhaar model of Daphnia base-line toxicity from MLOGP (mmol/l)	Molecular properties	1.144	0.138

Table 4.6 (Continued): Subset of 32 forward selected descriptors and sensitivity of MLP:32-2-1 to these descriptors

Descriptor	Definition	Group	Sensitivity*	SD**
Hypertens-80	Ghose-Viswanadhan-Wendoloski antihypertensive at 80% (drug-like index)	Molecular properties	1.129	0.071
BIC1	Bond information content (neighbourhood symmetry of 1-order)	Information indices	1.120	0.055
RDF055e	Radial Distribution Function - 5.5 / weighted by atomic Sanderson electronegativities	RDF descriptors	1.120	0.086
ATS6p	Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic polarizabilities	2D autocorrelations	1.104	0.094
MPC02	Molecular path count of order 02 (Gordon-Scantlebury index)	Walk and path counts	1.099	0.104
ACD/LogP	Log of octanol-water partition coefficient predicted by ACD Labs	Molecular properties	1.074	0.071
DP15	Molecular profile no. 15	Randic molecular profiles	1.063	0.096
p2Bp2B	Fragment containing two paths of length 2 with attached chains at position B	GSFRAG	1.050	0.098
Aeigv	Absolute eigenvalue sum from van der Waals weighted distance matrix	Eigenvalue-based indices	1.047	0.050
BEHm1	Highest eigenvalue n. 1 of Burden matrix / weighted by atomic masses	Burden Eigenvalues	1.047	0.033
GGI7	Topological charge index of order 7	Topological charge indices	1.040	0.058
D/D	Distance/detour index	Topological descriptors	1.040	0.042
N-068	Three aliphatic groups attached to a nitrogen atom	Atom-centred fragments	1.040	0.055
p4p4	Fragment containing two paths of length 4	GSFRAG	1.038	0.039
H7u	H autocorrelation of lag 7 / unweighted	GETAWAY descriptors	1.028	0.025
IDE	Mean information content on the distance equality	Information indices	1.025	0.025
RDF050e	Radial Distribution Function - 5.0 / weighted by atomic Sanderson electronegativities	RDF descriptors	1.024	0.039
H-048	H attached to C2(sp3) / C1(sp2) / C0(sp)	Atom-centred fragments	1.020	0.029
TIE	E-state topological parameter	Topological descriptors	1.019	0.022
Mor07p	3D-MoRSE - signal 07 / weighted by atomic polarizabilities	3D-MoRSE descriptors	1.018	0.015
p1p4-4N	Fragment containing two paths of length 1 and 4, with a nitrogen atom in position 4.	GSFRAG-L descriptors	1.018	0.020
H6v	H autocorrelation of lag 6 / weighted by atomic van der Waals volumes	GETAWAY descriptors	1.017	0.026
nR08	Number of 8-membered rings	Constitutional descriptors	1.015	0.005

Table 4.6 (Continued): Subset of 32 forward selected descriptors and sensitivity of MLP:32-2-1 to these descriptors

Descriptor	Definition	Group	Sensitivity*	SD**
Neoplastic-50	Ghose-Viswanadhan-Wendoloski antineoplastic at 50% (drug-like index)	Molecular properties	1.014	0.015
EEig11d	Eigenvalue 11 from edge adj. matrix weighted by dipole moments	Edge adjacency indices	1.013	0.006
R3v+	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes	GETAWAY descriptors	1.012	0.006
QXXe	Qxx COMMA2 value / weighted by atomic Sanderson electronegativities	Geometrical descriptors	1.009	0.003
p4-3N	Fragment containing a path of length 4, with a nitrogen atom in position 3.	GSFRAG-L descriptors	1.006	0.014
p1p1p2-6N	Fragment containing two paths of length 1 and one path of length 2, with a nitrogen atom in position 6.	GSFRAG-L descriptors	0.997	0.004

* Average ratio of network error without descriptor/with descriptor for n=6 networks, ** n=6.

4.3.3.3.2 Genetic feature selection for ANN retention time prediction

The genetic algorithm feature selection was run 10 times on the full 1096 descriptors and resulted in a different selection of descriptors each time, due to different randomly assigned initial populations. The only descriptors selected each time were ACD/LogP, ACD/LogD(pH=7), BLTD48, MLOGP2, ALOGP and the mean electrotopological state (Ms). The electrotopological state (E-state) of each skeletal atom in a molecule combines both the electronic character and the topological environment of the atom [139]. A unit penalty was used, which was multiplied by the number of variables in a network and added to the network selection error, to penalise networks containing large numbers of descriptors. Even when this unit penalty was set to 0.1 (compared to a default of 0.0001) the smallest number of descriptors selected was 181, still more than the number of compounds available to train the model. Therefore, further selection was required. Instead of running the genetic algorithm for further generations, the 'select a subset of variables' feature in the intelligent problem solver was used to develop a neural network using fewer molecular descriptors. The best model, MLP:11-3-1, used the eleven descriptors shown in Table 4.7. LogD was the only descriptor included in both the subset of forward selected and the subset of genetic algorithm selected descriptors. In both cases, the models were most sensitive to the removal of LogD and the network error was 1.552 times greater without LogD for the model using a subset of genetic algorithm selected descriptors.

Table 4.7: Subset of 11 genetic algorithm descriptors and sensitivity of MLP:11-3-1 to these descriptors

Descriptor	Definition	Group	Sensitivity*	SD**
LogD (pH = 7)	Log of the octanol water distribution coefficient at pH 7, predicted by ACD Labs.	Molecular properties	1.552	0.116
MLOGP2	Squared Moriguchi octanol-water partition coefficient	Molecular properties	1.282	0.140
RDF090E	Radial Distribution Function - 9.0 / weighted by atomic Sanderson electronegativities	RDF descriptors	1.155	0.114
HATS3P	Leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities	GETAWAY descriptors	1.129	0.004
ATS8P	Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic polarizabilities	2D autocorrelations	1.116	0.108
P3P5	Fragment containing one path of length 3 and one path of length 5	GSFRAG	1.057	0.066
BELE8	Lowest eigenvalue n. 8 of Burden matrix / weighted by atomic Sanderson electronegativities	Burden Eigenvalues	1.054	0.061
H-050	H attached to heteroatom	atom-centred fragments	1.051	0.021
JGI10	Mean topological charge index of order 10	Topological charge indices	1.043	0.058
P1P4B	Fragment containing one path of length 1 and one path of length 4 with attached chain at position B	GSFRAG	1.037	0.048
GATS6M	Geary autocorrelation - lag 6 / weighted by atomic masses	2D autocorrelations	1.008	0.013

* Average of n=6 networks, ** n=6

4.3.4 Performance of neural networks using selected descriptors

Correlation between measured and predicted retention times and prediction errors for the best neural networks using a subset of the descriptors selected by forward selection and genetic algorithms are shown in Figure 4.7. In both cases 3-layer multilayer perceptrons (MLPs) gave the best prediction models. The total range of prediction errors was larger (7.54 min compared to 6.59 min) for the model using 32 of the descriptors selected by forward selection (MLP:32-2-1) compared to the model using 11 of the descriptors selected by genetic algorithm (MLP:11-3-1), Figure 4.7 B, but this was still an improvement upon the smallest error range (8.45 min) seen for a model using molecular descriptors from a previous study (Figure 4.1 B(ii)). Further performance criteria are shown in Table 4.8 for MLP:32-2-1 and Table 4.9 for MLP:11-3-1. Overall, MLP:32-2-1 had a lower RMSE and MAE for the combined test set. However, the RMSE and MAE of the explosives test set were greater than the drugs test set for MLP:32-2-1 but lower than the than the drugs test set for MLP:11-3-1. Conversely, the RMSE and MAE were lower for the explosives training set than the drugs training set for MLP:32-2-1, but greater for MLP:11-3-1. As discussed earlier, compounds were randomly assigned to the training (n=105), select (n=22) and test (n=22)

Table 4.8: Performance of the best model and an ensemble of 6 models using a subset of 32 of the molecular descriptors selected by forward selection

Subset	Best model: MLP:32-2-1					Ensemble: 6xMLP:32-2-1				
	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)
Training - Explosives	0.92	0.97	0.13	1.12	0.95	0.92	0.92	0.51	1.16	0.99
Training - Drugs	0.83	0.81	1.49	1.59	1.31	0.84	0.78	1.85	1.60	1.29
Training - Combined	0.86	0.86	1.04	1.45	1.18	0.86	0.82	1.42	1.47	1.19
Select - Explosives	0.90	0.86	1.14	1.54	1.39	0.90	0.87	0.75	1.61	1.49
Select - Drugs	0.96	0.89	1.36	1.05	0.91	0.97	0.84	1.82	1.05	0.88
Select - Combined	0.94	0.87	1.36	1.20	1.04	0.94	0.84	1.60	1.23	1.05
Test - Explosives	0.87	0.99	-0.52	1.64	1.40	0.86	1.06	-1.04	1.74	1.60
Test - Drugs	0.94	0.94	0.43	0.92	0.70	0.93	0.85	1.24	1.03	0.84
Test - Combined	0.91	0.95	0.17	1.20	0.92	0.89	0.92	0.53	1.30	1.08
All	0.88	0.88	1.00	1.38	1.12	0.88	0.84	1.34	1.41	1.15

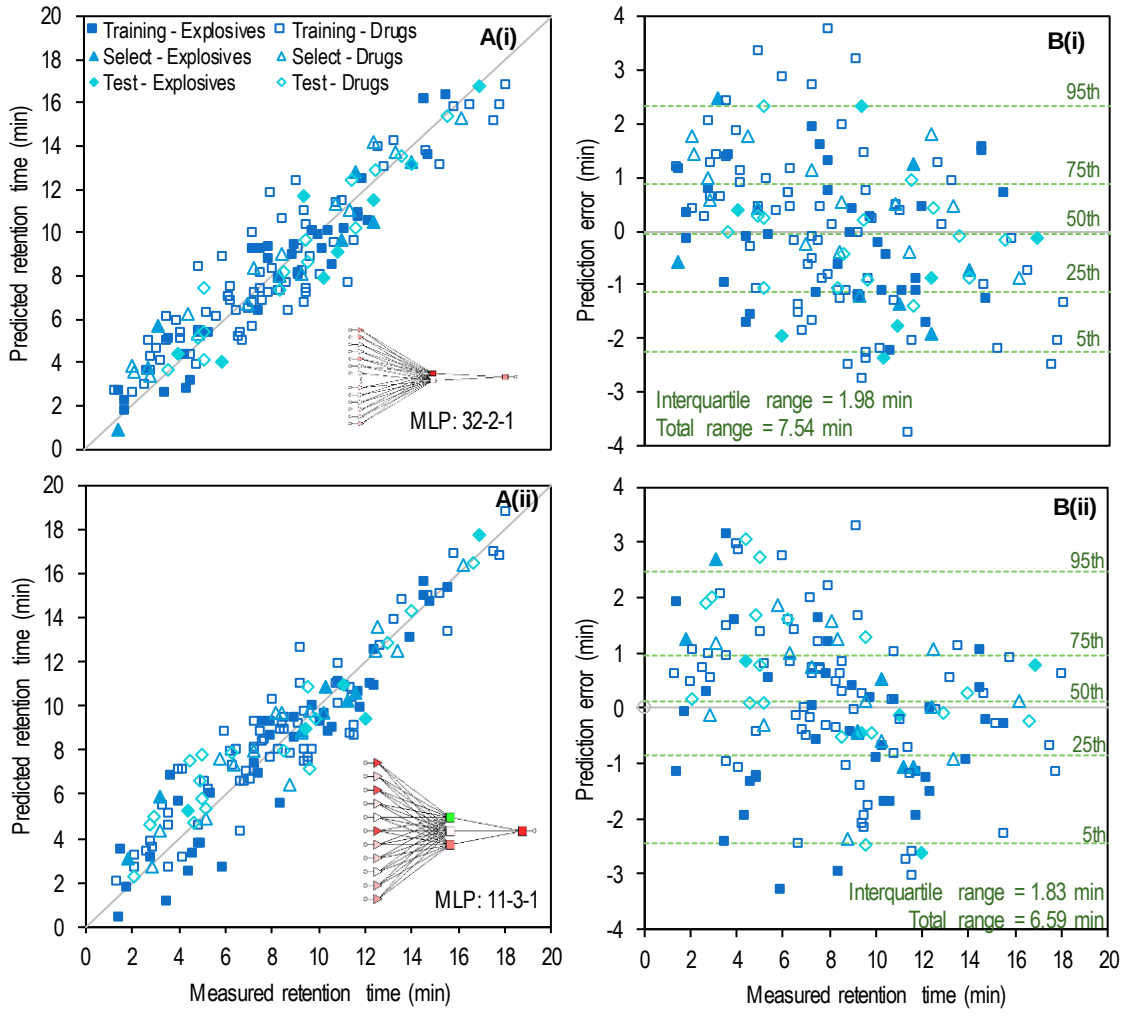


Figure 4.7: Correlation between measured and predicted retention times (A) and prediction error (B) for the best models using a subset of descriptors selected by (i) forward selection and (ii) genetic algorithm. Grey lines indicate perfect prediction and dashed green lines show percentiles.

Table 4.9: Performance of best model and an ensemble of six models using a subset of the descriptors selected by genetic selection.

Subset	Best model: MLP:11-3-1					Ensemble: 6xMLP:11-3-1				
	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)	R^2	Slope	Intercept (min)	RMSE (min)	MAE (min)
Training - Explosives	0.89	0.96	-0.08	1.44	1.14	0.89	0.82	1.21	1.44	1.15
Training - Drugs	0.87	0.88	1.15	1.41	1.14	0.83	0.80	1.79	1.62	1.34
Training - Combined	0.87	0.91	0.72	1.42	1.13	0.85	0.80	1.59	1.56	1.27
Select - Explosives	0.94	0.70	2.65	1.38	1.16	0.91	0.61	3.42	1.73	1.42
Select - Drugs	0.91	0.93	0.90	1.08	0.86	0.86	0.82	1.85	1.35	1.06
Select - Combined	0.91	0.86	1.48	1.17	0.94	0.86	0.75	2.36	1.46	1.16
Test - Explosives	0.91	0.96	0.11	1.29	0.96	0.83	0.90	0.53	1.77	1.60
Test - Drugs	0.90	0.84	1.88	1.50	1.14	0.87	0.78	2.19	1.63	1.27
Test - Combined	0.90	0.85	1.70	1.46	1.10	0.87	0.80	2.01	1.67	1.34
All	0.88	0.89	0.99	1.39	1.10	0.85	0.80	1.77	1.56	1.27

subsets during the first stage of network optimisation. Therefore, the 22 compounds included in the blind test set were not the same for both models; with 7 explosives in the test set for MLP:32-2-1 but only 5 explosives in the test set for MLP:11-3-1, and so the two subsets were not directly comparable. Performance of the two models for all subsets (training, selection and test) combined was very similar, with R^2=0.88, a slope of 0.88 or 0.89, an intercept of 1.00 or 0.99 min, a RMSE of 1.38 or 1.39 min and a MAE of 1.12 or 1.10 min. A paired T-test showed no significant difference (p=0.783) between the absolute errors of the two models.

4.3.4.1 Performance of ensembles

Ensembles of six repeated networks were also created, since the ultimate aim of developing a prediction model was to be able to generalise the model for the prediction of unknowns. Forming an ensemble of neural networks has been shown to improve generalisability [140]. Hansen and Salamon argued that forming an ensemble reduced generalisation errors which differ between individual networks, due to multiple local minima. Here, each of the six networks were given equal weightings and the retention time prediction from each model was averaged and the results are shown in Figure 4.8. For MLP:32-2-1, an ensemble of 6 networks with the same architecture reduced the total range of errors from 7.54 min (-3.79 to 3.75 min) to 6.90 min (-3.60 to 3.30 min). However, for MLP:11-3-1, the total error range of the ensemble (7.52 min) was larger than for the best individual network (6.59 min). The test sets should provide the best indication of generalisability, since compounds in the test set were not considered during training and

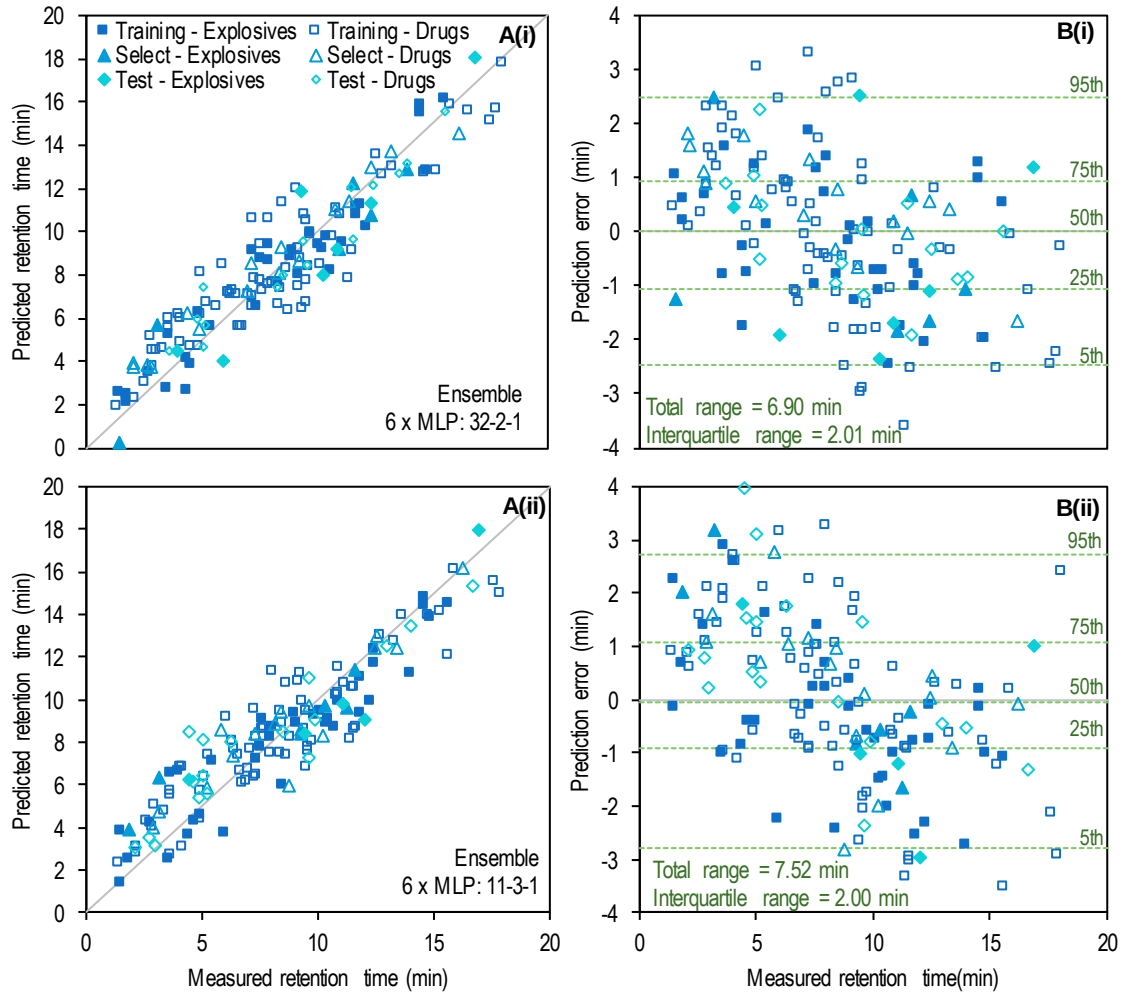


Figure 4.8: Correlation between measured and predicted retention times (A) and prediction error (B) for the ensembles (n=6) using a subset of descriptors selected by (i) forward selection and (ii) genetic algorithm. Grey lines indicate perfect prediction and dashed green lines show percentiles.

optimisation of the networks. However, as shown in Table 4.8 and Table 4.9, the RMSE and MAE for the explosives, drugs and combined test sets were larger for the ensemble than the best network for both MLP:32-2-1 and MLP:11-3-1. A paired T-test showed no significant difference between the absolute errors from the best model and ensemble of 6 models for MLP:32-2-1 ($p=0.388$), but a significant difference was seen for MLP:11-3-1 ($p=0.005$), with a larger MAE for the ensemble. As a result of this, and due to the narrower total error ranges, the best ensemble (6xMLP:32-2-1) and the best single network (MLP:11-3-1) were selected to examine rank prediction order.

4.3.4.2 Prediction order and rank correlation

In some instances, such as for the prediction of mannitol pentanitrate isomer elution order after three peaks have been detected, the ability to predict retention order may be just as valuable as predictive accuracy. Especially given the relatively large prediction errors in comparison to the differences in isomer retention times. Therefore, rank order correlation was also assessed. As shown in Table 4.10, Spearman's rank order correlation coefficient, r_s, was found to be statistically significant ($p<0.001$) for all models, suggesting that they were able to predict the general retention order. However, as shown in Figure 4.9 A, there was some scatter, especially in the middle of the run and in many of the 149 cases the retention order was incorrectly predicted. To investigate whether the retention order of a smaller number of more closely related compounds could be more accurately predicted, the retention order of the 12 nitrate esters included in the training, selection or test sets was also investigated. For both models, Spearman's correlation coefficient was greater for the nitrate esters alone (r_s 0.96 and 0.98) than for all compounds (r_s 0.92 and 0.93). As shown in Figure 4.9 B(i), the ensemble of 6xMLP:32-2-1 also correctly predicted the retention order of all the nitrate esters included in the training set; however, the retention orders of those included in the selection set (ETN and PETN) and test set (1,3-DNG) were incorrectly predicted. MLP:11-3-1 did correctly predict the retention order of ETN, which was again in the selection set (Figure 4.9 B(ii)); however, using this model the retention orders of two of the compounds in the training set (TEGDN and NG) were incorrectly predicted. Therefore, whilst there was strong correlation between the measured and predicted retention order, further work would be required (including the assessment of a greater number of related blind test compounds) before either model could be used to confidently predict retention order for new/unknown compounds.

Table 4.10: Spearman's rank order correlation between predicted and measured retention times

	All compounds (n=149)		Nitrate esters only (n=12)	
Neural network	Spearman's coefficient (r_s)	Significance (p)	Spearman's coefficient (r_s)	Significance (p)
6xMLP:32-2-1	0.926	<0.001	0.963	<0.001
MLP:11-3-1	0.922	<0.001	0.984	<0.001

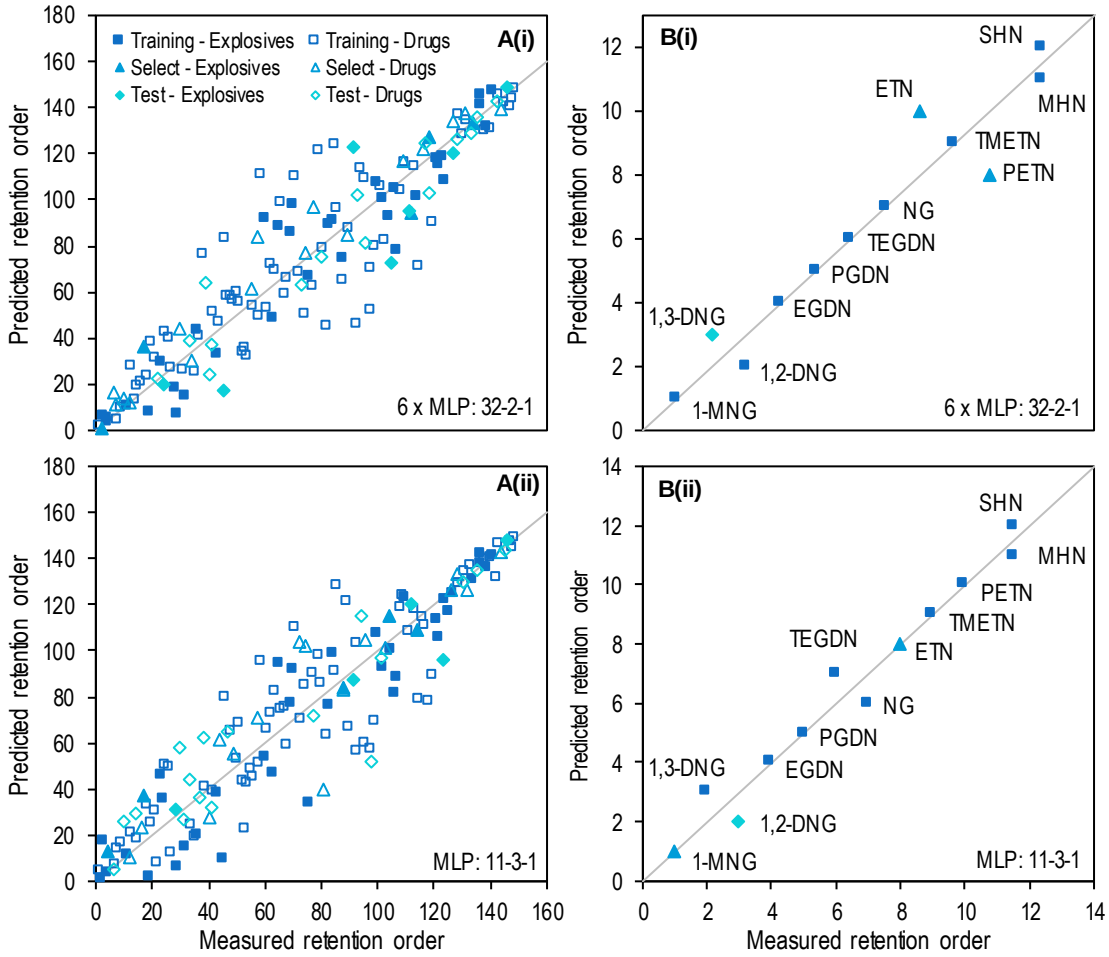


Figure 4.9: Rank order correlation for A) all 149 compounds and B) nitrate esters using I) the ensemble of 6 MLP:32-2-1 and ii) the best MLP:11-3-1.

4.3.4.3 Setting a retention range for prediction of new compounds

When predicting the retention time of new/unknown compounds it is the individual prediction error rather than the mean prediction error that is of interest. The prediction interval (PI) is the range likely to include the response of a single new observation and so this was calculated instead of the confidence interval (CI) of prediction, the range likely to contain the mean response. Due to the increased uncertainty of predicting a single response rather than the mean response, the PI is normally wider than the CI. The prediction interval was given by the new predicted value ($\hat{y}_h$) plus or minus the appropriate t-value (*t*) multiplied by the standard error of prediction (s_{pred}) [141], as shown in Equation 4.1. As the standard error of prediction for the entire population was not known, an estimate of the standard error of prediction (s_{est}), was calculated based on the standard error of the sample of 149 compounds used for training and testing the networks using Equation 4.2 [142], where *Y* and *Y'* were known measured and predicted retention time pairs and *N* was

the number of compounds in the sample. For the sample size used here (149 compounds), the t-value (to 3 significant figures) was 1.98 for a 95% PI, close to the t-value for infinity degrees of freedom (1.96) [143,144]. The calculated prediction intervals are shown in Table 4.11. As expected, the size of the prediction intervals increased as the confidence level increased. For a 99 % probability that the true retention time would fall within the prediction interval, a 7.4 min (± 3.7 min) PI was required. Whereas, if only a 75 % probability that the true retention time would fall within the PI was required, a 3.2 min (± 1.6 min) PI could be used. In order to use the prediction model for exclusion purposes, the highest possible probability that the true value would fall within the PI should be used to limit false negatives. However, when prioritising compounds to obtain reference materials for confirmation of unknowns a more relaxed PI might be more appropriate, at least in the first instance, to limit the number of compounds for which reference materials need to be purchased and/or synthesised.

Equation 4.1: Prediction interval

$$PI = \hat{y}_h \pm t_{(\alpha/2,n-2)} \times s_{pred}$$

Equation 4.2: Standard error of prediction

$$s_{est} = \sqrt{\frac{\sum(Y - Y')^2}{N - 2}}$$

Table 4.11: Prediction intervals

	Prediction interval (min)					
Model	50 %	75 %	80 %	90 %	95 %	99 %
MLP:11-3-1	± 0.95	± 1.62	± 1.81	± 2.32	± 2.77	± 3.67
6xMLP:11-3-1	± 1.06	± 1.82	± 2.03	± 2.61	± 3.11	± 4.11
MLP:32-2-1	± 0.94	± 1.61	± 1.79	± 2.31	± 2.75	± 3.64
6xMLP:32-2-1	± 0.96	± 1.64	± 1.83	± 2.36	± 2.81	± 3.72

Calculation of the PI was based on the assumption that the residual errors were normally distributed; since for a normal distribution, 95 % of responses fall within 2 standard deviations of the population mean. Although, as neither the population mean, nor the population standard deviation were known in this case, they were estimated based on the sample and the errors associated with using such an estimate. As shown by the roughly bell-shaped histograms (Figure 4.10 A) and the minimal deviation from the line in the normal Q-Q plots (Figure 4.10 B), the residual errors were approximately normally distributed, with the mean error close to zero, for both the ensemble of 6 MLP:32-2-1 and MLP:11-3-1. Additionally, the Shapiro-Wilk test for

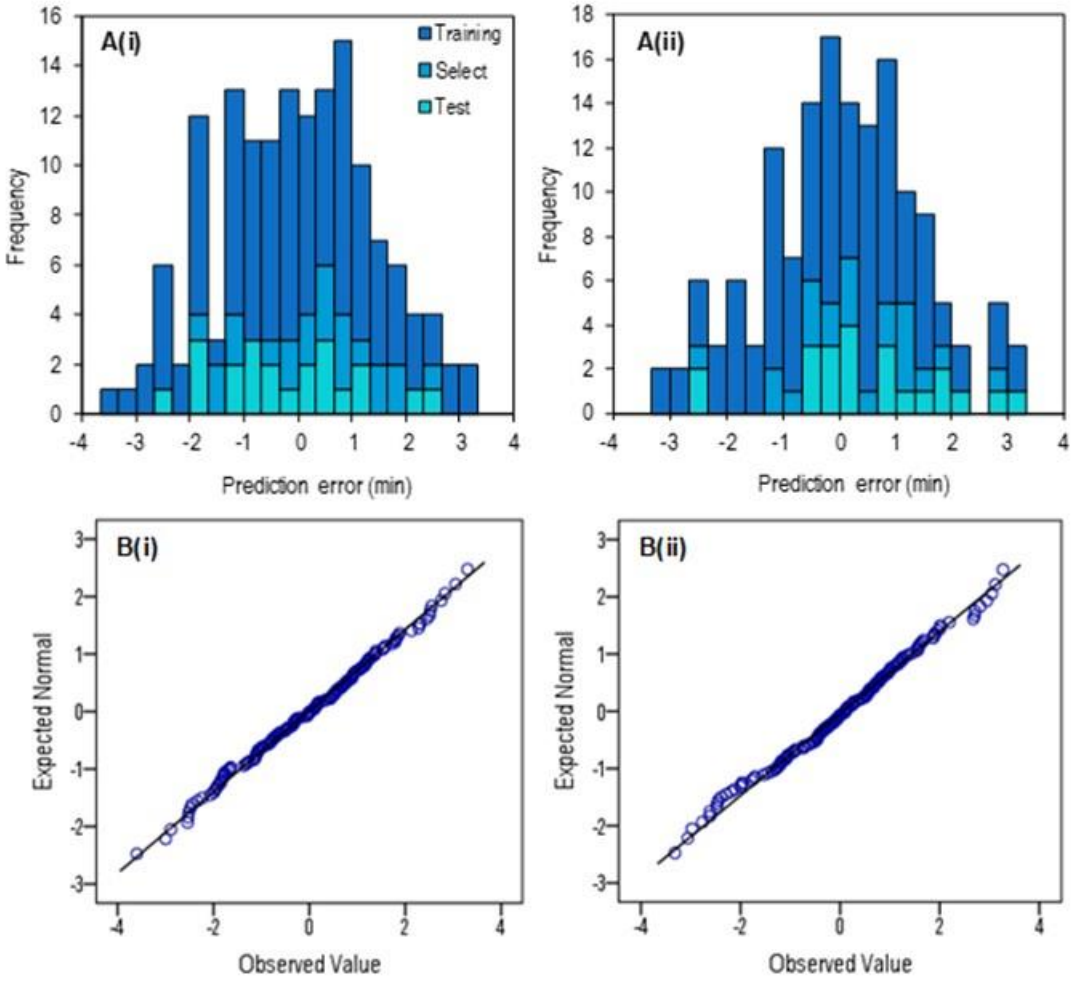


Figure 4.10: Assessment of normal distribution of prediction errors using A) histograms and B) normal Q-Q plots for i) 6xMLP:33-2-1 and ii) MLP:11-3-1.

normality showed no significant deviation from normality for either the ensemble of 6 MLP:32-2-1 (p=0.869, 0.684 and 0.530 for training, select and test sets respectively) or MLP:11-3-1 (p=0.692, 0.976 and 0.359 for training, select and test sets respectively).

4.3.4.4 Summary of performance assessment

Based on the assessment of performance presented in this section, it was not possible to conclusively determine the best model for predicting the retention times of new compounds. The 99 % prediction intervals for MLP:11-3-1, MLP:32-2-1 and the ensemble 6xMLP:32-2-1 were very similar (± 3.67, 3.64 and 3.72 min, respectively). As were the overall MAE (1.10, 1.12 and 1.15 min, respectively) and the RMSE (1.39, 1.38 and 1.41 min respectively). Therefore, further assessment of method performance, preferably for new compounds not considered during training, optimisation or selection of the neural networks, was necessary before one model could be selected over the others.

4.3.5 Application of retention time prediction to new compounds

4.3.5.1 Prediction of organic gunshot residue retention times

Several of the energetic materials included in the explosives dataset may also be found in organic gunshot residue (OGSR). For example, DPA, DEDPU, DMDPU, NG and 2,4-DNT have all been detected in smokeless powders [39]. A gun surveillance standard including dimethyl phthalate (DMP), 2,4'-, 2,4-, 2,2'- and 4,4'-dinitrodiphenylamine (DNDPA), 2- and 4-nitrodiphenylamine (NDPA), N-Nitrosodiphenylamine (N-NDPA) and diphenylamine (DPA, also included in explosives dataset) was analysed on the same day as the mixture used to develop the ANNs. The extracted ion chromatograms are shown in Figure 4.11. Multiple peaks were detected in several of the EICs due to DNDPA and NDPA isomers and closely related compounds N-NDPA and DPA which produced common ions. In all cases there was baseline separation, highlighting

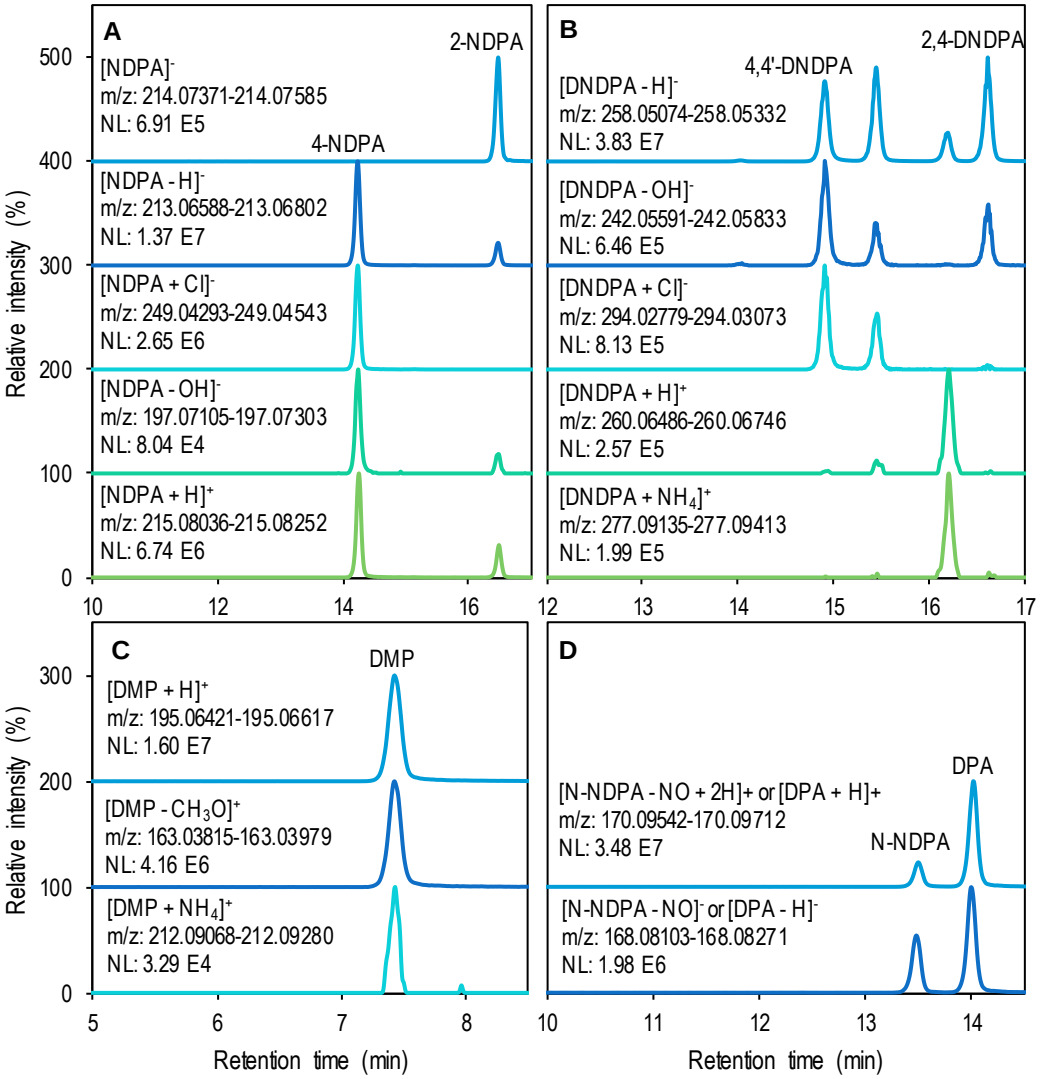


Figure 4.11: EICs from gun surveillance standard for A) NDPAs, B) DNDPAs, C) DMP and D) N-NDPA.

the value of optimising chromatography during Chapter 2. At least three ions, including both a molecular ion or (de)protonated molecule and a fragment ion, were detected for all compounds apart from N-NDPA, reinforcing the generalisability of the developed LC-HRMS method. In the case of N-NDPA, two ions were still detected but in both cases the nitroso (-N=O) group was lost during in-source fragmentation. As seen for other nitroaromatics, the nitro- and dinitro-DPAs produced $[M\text{-}OH]^-$ fragment ions and ion ratios differed between isomers.

Individual standards were also analysed for DMP, 2,4- and 4,4'-DNDPA, 2- and 4- NDPA and N-NDPA to confirm retention times and order (labelled in Figure 4.11). These six compounds were then used as an external test set to further assess performance of the prediction models and their chemical structures are shown in Figure 4.12. Without reference standards it was not possible to identify which of the middle two DNDPA peaks were due to 2,2'- or 2,4'-DNDPA and so these compounds were not included in the external test set.

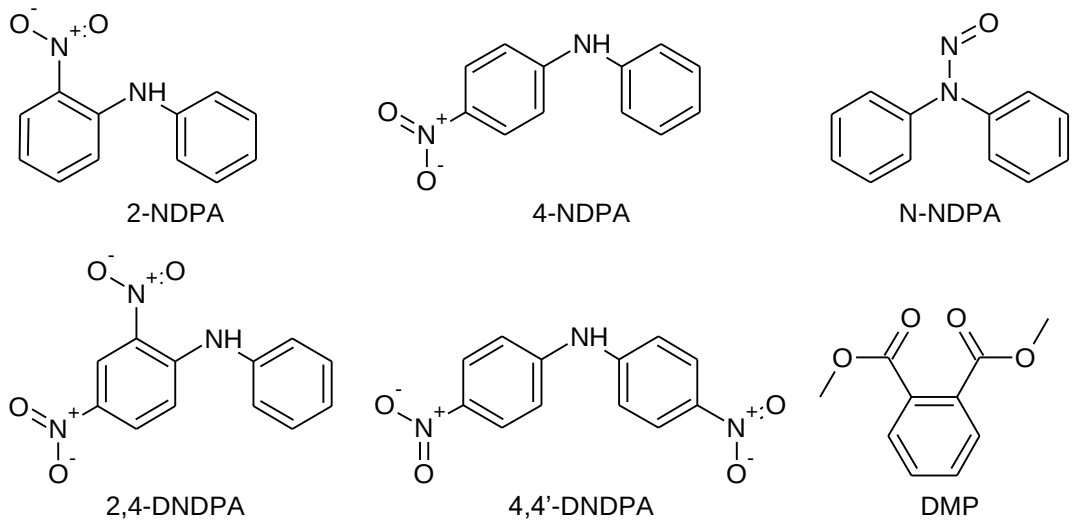


Figure 4.12: Chemical structures of new OGSR compounds used as an external test set for validation of prediction models.

The predicted retention times and prediction errors for the six new, to the prediction model at least, energetic materials, using MLP:11-3-1, MLP:32-2-1 and 6xMLP:32-2-1, are shown in Table 4.12. This small external test set indicated that MLP:32-2-1 and 6XMLP:32-2-1 were more generalisable and accurate than MLP:11-3-1. All predicted retention times using MLP:32-2-1 and 6xMLP:32-2-1 fell within the 95 % PIs (± 2.75 and 2.81 min, respectively) and 4 out of 6 fell within the 50 % PI (± 0.94 and 0.96 min, respectively). In contrast, 3 out of 6 predicted retention times using MLP:11-3-1 fell outside the 99 % PI (± 3.67 min). This demonstrated the importance of

Table 4.12: Measured and predicted retention times for OGSR components.

OGSR	Ion	Exact *m/z*	t_R^M (min)	SD (min)	MLP:11-3-1 t_R^P (min)	MLP:11-3-1 Δt_R (min)	MLP:32-2-1 t_R^P (min)	MLP:32-2-1 Δt_R (min)	6xMLP:32-2-1 t_R^P (min)	6xMLP:32-2-1 Δt_R (min)
DMP	$[M+H]^+$	195.0652	7.38	0.02	9.29	1.91	7.91	0.53	7.50	0.12
N-NDPA	$[M-NO]^-$	168.0819	13.45	0.02	12.31	-1.14	14.27	0.82	14.04	0.59
4-NDPA	$[M-H]^-$	213.0670	14.19	0.03	10.66	-3.53	14.53	0.34	14.22	0.03
4,4'-DNDPA	$[M-H]^-$	258.0520	14.87	0.03	10.39	-4.48	14.05	-0.82	13.97	-0.90
2-NDPA	$[M-H]^-$	213.0670	16.43	0.02	12.44	-3.99	14.76	-1.67	14.37	-2.06
2,4-DNDPA	$[M-H]^-$	258.0520	16.58	0.03	11.72	-4.86	13.94	-2.64	13.87	-2.71
			MAE (min)		3.32		1.14		1.07	
			RMSE (min)		3.58		1.38		1.46	

validating prediction models using an external test set with known retention times, before relying on a prediction model to make decisions. In terms of retention order, all three models correctly predicted that 4-NDPA would elute before 2-NDPA. MLP:11-3-1 also correctly predicted that 4,4'-DNDPA would elute before 2,4-DNDPA, but MLP:32-2-1 and 6xMLP:32-2-1 did not. MLP:32-2-1 and 6xMLP:32-2-1, also underestimated the separation achieved between the two pairs of closely related structural isomers. However, overall prediction accuracy was better for MLP:32-2-1 and 6xMLP:32-2-1, with MAE and RMSE less than half those for MLP:11-3-1. Additionally, the MAE and RMSE using MLP:11-3-1 were 3.0 and 2.6 times greater, respectively, for the new compounds than for the combined training, selection and test sets used to develop and select the model. Whereas, using MLP:32-2-1 and 6xMLP:32-2-1 the MAE and RMSE for the new compounds were within 0.05 min of the MAE and RMSE for the original dataset.

Even this relatively small validation set called into question the generalisability of the MLP:11-3-1 model and the prediction intervals calculated in Section 4.3.4.3 for this network. However, the MLP:32-2-1 single network, 6XMLP:32-2-1 ensemble and their PIs performed better. Prediction errors for the MLP:32-2-1 single network and 6xMLP:32-2-1 ensemble were similar, with the MAE slightly lower for the ensemble and the RMSE slightly lower for the single network. Based on previous work suggesting that the formation of ensembles improves generalisability [140], 6xMLP:32-2-1 was selected to use for retention time prediction of MEKP oligomers, sugar nitrates and structural isomers of RDX, TATB and PYX. Prediction of these compounds using MLP:11-32-1 and MLP:32-2-1 can be found in the Appendix (Table A. 5 and Table A. 7). Further work

may also be required to investigate generalisability and applicability of the PIs to a wider variety of new compounds.

4.3.5.2 Prediction of MEKP oligomer retention times

As discussed in Chapter 3, Section 3.3.2.4, Methyl ethyl ketone peroxide (MEKP) includes cyclic peroxide (CP) and linear dihydroperoxy peroxide (DHP) oligomers of various sizes (Figure 3.9) and reference materials for individual oligomers were not available. A number of these oligomers were detected in a sample of MEKP analysed using this LC-HRMS method, but it was not always clear which chromatographic peak corresponded to which oligomer due to multiple peaks in some extracted ion chromatograms and peaks with the same retention time seen in multiple EICs (Figure 3.10). Here, MEKP was revisited again to see if the predictive model (6xMLP:32-2-1) could aid assignment of chromatographic peaks to specific MEKP oligomers. The MEKP sample was analysed again on the same day as the ANN training, selection and test compounds and the measured retention times (t_R^M) obtained for the EIC corresponding to the ammonium adduct of each oligomer are shown in Table 4.13. Fewer peaks were detected than the first time the MEKP sample was analysed and no peaks were detected in the EIC for the CP tetramer this time. This may have been due to sample degradation, since several months elapsed in between analysis.

The predicted retention times (t_R^P) are also shown in Table 4.13, along with the prediction order and the difference in retention time from each of the chromatographic peaks detected. All retention times fell within the 99 % prediction interval, further supporting the generalisability of the ensemble, 6XMLP:32-2-1. As expected, the prediction model ordered the oligomers by size with

Table 4.13: Predicted retention time and order for MEKP oligomers using 6xMLP:32-2-1.

MEKP Oligomer	Formula	$[M+NH4]^+$ Exact *m/z*	t_R^M (min)	t_R^P (min)	t_R^P Order	Δt_R (min)
CP dimer	$C_8H_{16}O_4$	194.1387	8.66 8.82	10.29	1	1.63 1.47
DHP dimer	$C_8H_{18}O_6$	228.1442	8.66 8.77	12.11	2	3.45 3.34
DHP trimer	$C_{12}H_{26}O_8$	316.1966	12.93	14.62	3	1.69
CP trimer	$C_{12}H_{24}O_6$	282.1911	17.14	14.72	4	-2.42
CP tetramer	$C_{16}H_{32}O_8$	370.2435	n.d.*	16.14	5	-
DHP tetramer	$C_{16}H_{34}O_{10}$	404.2490	16.01	17.19	6	1.18
DHP pentamer	$C_{20}H_{42}O_{12}$	492.3015	18.14 18.29	17.64	7	-0.5 -0.65

t_R^M – measured retention time for peaks detected with *m/z* ± 5 ppm of $[M+NH_4]^+$ ion; t_R^P – Predicted retention time; Δt_R – difference between prediction and measured retention time; n.d. not detected; *Previously detected with a retention time of 18.23 min.

dimer < trimer < tetramer < pentamer. What was perhaps more surprising was that while the predicted order was DHP < CP for trimers, it was CP < DHP for the dimers and tetramers. Due to the polar dihydroperoxy (R-O-OH) groups at each end of the DHP oligomers, they might be expected to be more polar than the corresponding CP oligomer. This was reflected in the ALOGP and MLOGP predicted LogP values where DHP < CP for the dimer, trimer and tetramer (Table 4.14). However, for ACD/LogP this trend was reversed with CP < DHP for all three oligomer pairs. While ACD/LogP was the only predicted LogP included in the 6xMLP:32-2-1 model; BLTD48 another descriptor used by the model was highly correlated (R^2> 0.9999) with MLOGP which, along with contributions from the other 30 descriptors, may explain the switch in prediction order for this model. Based on the prediction order of the model, the CP dimer would elute first (i.e. 8.66 min) and the DHP dimer would elute second (i.e. 8.77 or 8.82 min). However, both peaks were closer to the predicted retention time of the CP dimer and there was significant overlap between the prediction intervals. Therefore, if only one peak was detected the model would not be able to accurately predict whether that peak was due to the CP or DHP dimer.

Table 4.14: LogP values and order for MEKP oligomers

MEKP Oligomer	t_R^M (min)	ACD/LogP	ACD/LogP Order	ALOGP	ALOGP Order	MLOGP	MLOGP Order
CP dimer	8.66 8.82	2.08	1	3.217	2	1.793	2
DHP dimer	8.66 8.77	3.19	2	2.297	1	1.438	1
DHP trimer	12.93	4.91	5	3.906	3	2.341	3
CP trimer	17.14	3.44	3	4.826	4	2.645	4
CP tetramer	n.d.*	4.5	4	6.435	6	3.396	6
DHP tetramer	16.01	6.65	6	5.514	5	3.131	5
DHP pentamer	18.14 18.29	8.28	7	7.123	7	3.849	7

t_R^M – Measured retention time for peaks detected with *m/z* ± 5 ppm of $[M+NH_4]^+$ ion;
n.d. not detected; *Previously detected with a retention time of 18.23 min.

4.3.5.3 Prediction of nitrated sugar retention times

Purity of the mannitol hexanitrate (MHN) and sorbitol hexanitrate (SHN) samples was unknown, as would be the case for real samples encountered during forensic casework. As shown in Chapter 3, both the MHN and SHN samples contained multiple chromatographic peaks (Figure 3.5) and it was proposed that these additional chromatographic peaks were due to less-nitrated

sugars (i.e. penta-, tetra-, tri-, di- or mono-nitrates). Multiple ions were also detected at each retention time and in the case of the peak at 13.73 min in the MHN sample, five ions were detected including ions with 3, 4 and 5 nitrogen atoms. While it was thought likely that the general retention order was mononitrates < dinitrates < trinitrate < tetranitrates < pentanitrates < hexanitrates; due to similarities in structure, it was not always clear which nitrated sugar was the source of each ion. For example, the ion with an elemental composition of $C_6H_9O_{14}N_4^-$ could be due to either a tetranitrate $[M-H]^-$ ion, a pentanitrate $[M-NO_2]^-$ ion or a hexanitrate $[M-2NO_2+H]^-$ ion (Figure 4.13). The $C_6H_9O_{14}N_4^-$ ion could have the same structure, regardless of whether it was formed by APCI of MHN, an MPN isomer or an MTN isomer. Therefore, in this instance, tandem mass spectrometry would not help identify the source of the ion.

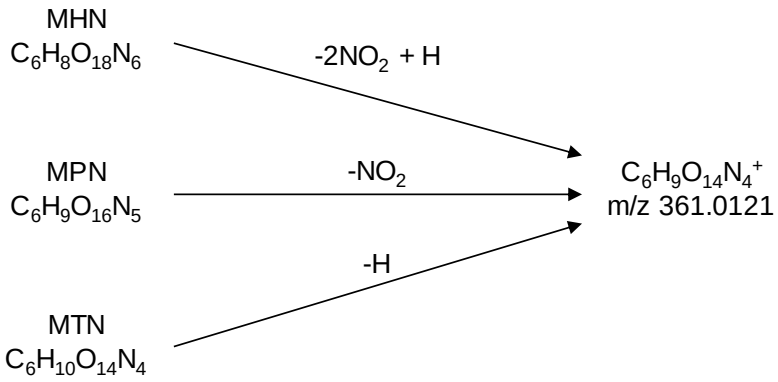


Figure 4.13: Formation of a $C_6H_9O_{14}N_4^+$ ion from mannitol hexanitrate (MHN), pentanitrate (MPN) and tetranitrate (MTN).

Here retention times were predicted for various mannitol nitrates using the ensemble of 6xMLP:32-2-1 (Table 4.15). The model did predict retention times for the hexanitrate (MHN) > the pentanitrates (MPNs) > the tetranitrates (MTN) > the trinitrates (MTriN), as expected due to the greater polarity and hydrogen bonding ability of hydroxyl (-OH) groups compared to nitrate ($-ONO_2$) groups. The predicted retention time for MHN was closest to the peak detected at 14.55 min (Δt_R = 1.06 min), supporting assignment of this peak as MHN. Although, the MTriN isomers were the only compounds to fall outside the 99 % PI for this peak and so the only compounds that could be excluded using a 99 % PI. The MPNs had the closest predicted retention times to the peaks at 13.73 and 13.29 min (Δt_R = 0.40-0.70 and 0.04-0.26 min, respectively). Again, the MTriN isomers fell outside the 99 % PI for these peaks, supporting the theory that the $C_6H_{10}O_{12}N_3^-$ ions were fragment ions from MPN or MTN isomers, rather than deprotonated MTriN molecules.

Table 4.15: Predicted retention times for mannitol nitrates using ensemble of 6xMLP:32-2-1

	t_R^M (min)	14.55	13.73	13.29	12.52	10.83	10.41	9.98	9.83	9.25	
	Ions detected	$C_6H_8O_{21}N_7^-$ $C_6H_8O_{16}N_5^-$ $C_6H_9O_{16}N_5Cl^-$ $C_6H_9O_{14}N_4^-$	$C_6H_8O_{16}N_5^-$ $C_6H_9O_{16}N_5Cl^-$ $C_6H_9O_{14}N_4^-$ $C_6H_{10}O_{14}N_4Cl^-$ $C_6H_{10}O_{12}N_3^-$	$C_6H_9O_{16}N_5Cl^-$ $C_6H_9O_{14}N_4^-$ $C_6H_{10}O_{14}N_4Cl^-$	$C_6H_9O_{16}N_5Cl^-$ $C_6H_9O_{14}N_4^-$	$C_6H_9O_{14}N_4^-$ $C_6H_{10}O_{14}N_4Cl^-$ $C_6H_{10}O_{12}N_3^-$	$C_6H_{10}O_{14}N_4Cl^-$ $C_6H_{10}O_{12}N_3^-$	$C_6H_{10}O_{14}N_4Cl^-$ $C_6H_{10}O_{12}N_3^-$	$C_6H_9O_{14}N_4^-$ $C_6H_{10}O_{14}N_4Cl^-$ $C_6H_{10}O_{12}N_3^-$	$C_6H_{10}O_{14}N_4Cl^-$ $C_6H_{10}O_{12}N_3^-$	
Compound	t_R^P (min)	Δt_R	Δt_R	Δt_R	Δt_R	Δt_R	Δt_R	Δt_R	Δt_R	Δt_R	
MHN	15.61	1.06	1.88	2.32	3.09	~~4.78~~	~~5.20~~	~~5.63~~	~~5.78~~	~~6.36~~	Key*
M-1,2,3,5,6-PN	13.33	1.22	0.40	0.04	0.81	2.50	2.92	3.35	3.50	~~4.08~~	25 % PI
M-1,2,3,4,5-PN	13.06	1.49	0.67	0.23	0.54	2.23	2.65	3.08	3.23	~~3.81~~	50 % PI
M-1,2,3,4,6-PN	13.03	1.52	0.70	0.26	0.51	2.20	2.62	3.05	3.20	~~3.78~~	75 % PI
M-1,2,4,5-TN	12.01	2.54	1.72	1.28	0.51	1.18	1.60	2.03	2.18	2.76	80 % PI
M-1,3,4,6-TN	11.91	2.64	1.82	1.38	0.61	1.08	1.50	1.93	2.08	2.66	90 % PI
M-2,3,4,5-TN	11.72	2.83	2.01	1.57	0.80	0.89	1.31	1.74	1.89	2.47	95 % PI
M-1,2,4,6-TN	11.69	2.86	2.04	1.60	0.83	0.86	1.28	1.71	1.86	2.44	99 % PI
M-1,3,4,5-TN	11.67	2.88	2.06	1.62	0.85	0.84	1.26	1.69	1.84	2.42	
M-1,2,3,6-TN	11.66	2.89	2.07	1.63	0.86	0.83	1.25	1.68	1.83	2.41	
M-1,2,5,6-TN	11.53	3.02	2.20	1.76	0.99	0.70	1.12	1.55	1.70	2.28	
M-1,2,3,4-TN	11.39	3.16	2.34	1.90	1.13	0.56	0.98	1.41	1.56	2.14	
M-1,2,3,5-TN	11.26	3.29	2.47	2.03	1.26	0.43	0.85	1.28	1.43	2.01	
M-1,3,6-TriN	7.96	~~6.59~~	~~5.77~~	~~5.33~~	~~4.56~~	2.87	2.45	2.02	1.87	1.29	
M-1,3,4-TriN	7.65	~~6.90~~	~~6.08~~	~~5.64~~	~~4.87~~	3.18	2.76	2.33	2.18	1.60	
M-1,2,6-TriN	7.63	~~6.92~~	~~6.10~~	~~5.66~~	~~4.89~~	3.20	2.78	2.35	2.20	1.62	
M-1,2,3-TriN	7.63	~~6.92~~	~~6.10~~	~~5.66~~	~~4.89~~	3.20	2.78	2.35	2.20	1.62	
M-2,3,4-TriN	7.53	~~7.02~~	~~6.20~~	~~5.76~~	~~4.99~~	3.30	2.88	2.45	2.30	1.72	
M-1,2,4-TriN	7.52	~~7.04~~	~~6.22~~	~~5.78~~	~~5.01~~	3.32	2.90	2.47	2.32	1.74	

*Δt_R values highlighted in white – dark green depending on the PI, with darker green indicating values closer to measured t_R. Values outside of the 99 % PI shown in red and struck through

Retention times were also predicted for sorbitol nitrates using the ensemble of 6xMLP:32-2-1 (Table 4.16). As sorbitol has less symmetry than mannitol, there are more isomers for the sorbitol nitrates than the mannitol nitrates. Additionally, ions with a lower *m/z* and fewer nitrogen atoms were detected in the SHN sample, suggesting the presence of less-nitrated sugars, i.e. STriN and SDN as well as STN, SPN and SHN (Table 4.17). As with the mannitol nitrates, the predicted retention times followed the general trend of predicted retention times for sorbitol hexanitrate (SHN) > the pentanitrate (SPN) isomers > the tetranitrate (STN) isomers > the trinitrate (STriN) isomers > the dinitrate (SDN) isomers > the mononitrate (SMN) isomers. SHN, all SPN isomers and all STN isomers had predicted retention times outside of the 99 % PI for the peaks at 6.45, 5.02, 4.88, 3.75 and 2.36 min, supporting the theory that the ions detected at these retention times (Table 4.17) were due to STriN or SDN isomers rather than fragment ions from the more-nitrated sugars. For the peak at 2.36 min, the retention times predicted for the STriN isomers also fell outside the 99 % PI suggesting that the $C_6H_{11}O_{10}N_2^-$ ion detected at this retention time was due to a SDN $[M-H]^-$ rather than a STriN $[M-NO_2]^-$.

It is worth pointing out that if the MLP:11-3-1 network had been selected to generate predicted retention times then the predicted retention times and conclusions made based on prediction intervals would have differed. For example, using MLP:11-3-1, predicted retention times for some STN isomers did fall within the 99% PI for the peaks at 6.45, 5.02 and 4.88 min and the predicted retention time for some STriN isomers fell within the 99 % PI for the peak at 2.36 min (Table A. 5). This highlights the risk associated with relying upon predicted retention times for identification and the importance of selecting the best model for prediction, since different models will lead to different results. Ideally, predicted retention times should later be confirmed using reference standards, with predicted retention times used to prioritise the compounds for which reference materials are purchased or synthesised. However, here individual reference standards were not available for the sugar nitrates and so this was not possible.

Table 4.16: Predicted retention times for sorbitol nitrates using the ensemble of 6xMLP:32-2-1

Compound	t_R^M (min) t_R^P (min)	14.50 Δt_R	13.37 Δt_R	12.91 Δt_R	12.62 Δt_R	10.30 Δt_R	10.07 Δt_R	9.77 Δt_R	9.24 Δt_R	8.84 Δt_R	6.45 Δt_R	5.02 Δt_R	4.88 Δt_R	3.75 Δt_R	2.36 Δt_R	
SHN	15.84	1.34	2.47	2.93	3.22	~~5.54~~	~~5.77~~	~~6.07~~	~~6.60~~	~~7.00~~	~~9.39~~	~~10.82~~	~~10.96~~	~~12.09~~	~~13.48~~	Key*
S-1,2,3,5,6-PN	13.42	1.08	0.05	0.51	0.80	3.12	3.35	3.65	~~4.18~~	~~4.58~~	~~6.97~~	~~8.40~~	~~8.54~~	~~9.67~~	~~11.06~~	25 % PI
S-1,2,3,4,6-PN	13.40	1.10	0.03	0.49	0.78	3.10	3.33	3.63	~~4.16~~	~~4.56~~	~~6.95~~	~~8.38~~	~~8.52~~	~~9.65~~	~~11.04~~	50 % PI
S-1,2,4,5,6-PN	13.36	1.14	0.01	0.45	0.74	3.06	3.29	3.59	~~4.12~~	~~4.52~~	~~6.91~~	~~8.34~~	~~8.48~~	~~9.61~~	~~11.00~~	75 % PI
S-1,2,3,4,5-PN	13.21	1.29	0.16	0.30	0.59	2.91	3.14	3.44	~~3.97~~	~~4.37~~	~~6.76~~	~~8.19~~	~~8.33~~	~~9.46~~	~~10.85~~	80 % PI
S-1,3,4,5,6-PN	13.00	1.50	0.37	0.09	0.38	2.70	2.93	3.23	~~3.76~~	~~4.16~~	~~6.55~~	~~7.98~~	~~8.12~~	~~9.25~~	~~10.64~~	90 % PI
S-2,3,4,5,6-PN	12.90	1.60	0.47	0.01	0.28	2.60	2.83	3.13	3.66	~~4.06~~	~~6.45~~	~~7.88~~	~~8.02~~	~~9.15~~	~~10.54~~	95 % PI
S-2,3,5,6-TN	12.23	2.27	1.14	0.68	0.39	1.93	2.16	2.46	2.99	3.39	~~5.78~~	~~7.21~~	~~7.35~~	~~8.48~~	~~9.87~~	99 % PI
S-1,3,4,6-TN	11.98	2.52	1.39	0.93	0.64	1.68	1.91	2.21	2.74	3.14	~~5.53~~	~~6.96~~	~~7.10~~	~~8.23~~	~~9.62~~	
S-3,4,5,6-TN	11.89	2.61	1.48	1.02	0.73	1.59	1.82	2.12	2.65	3.05	~~5.44~~	~~6.87~~	~~7.01~~	~~8.14~~	~~9.53~~	
S-1,4,5,6-TN	11.84	2.66	1.53	1.07	0.78	1.54	1.77	2.07	2.60	3.00	~~5.39~~	~~6.82~~	~~6.96~~	~~8.09~~	~~9.48~~	
S-2,4,5,6-TN	11.77	2.73	1.60	1.14	0.85	1.47	1.70	2.00	2.53	2.93	~~5.32~~	~~6.75~~	~~6.89~~	~~8.02~~	~~9.41~~	
S-1,3,4,5-TN	11.76	2.74	1.61	1.15	0.86	1.46	1.69	1.99	2.52	2.92	~~5.31~~	~~6.74~~	~~6.88~~	~~8.01~~	~~9.40~~	
S-1,2,3,6-TN	11.75	2.75	1.62	1.16	0.87	1.45	1.68	1.98	2.51	2.91	~~5.30~~	~~6.73~~	~~6.87~~	~~8.00~~	~~9.39~~	
S-2,3,4,6-TN	11.61	2.89	1.76	1.30	1.01	1.31	1.54	1.84	2.37	2.77	~~5.16~~	~~6.59~~	~~6.73~~	~~7.86~~	~~9.25~~	
S-1,3,5,6-TN	11.59	2.91	1.78	1.32	1.03	1.29	1.52	1.82	2.35	2.75	~~5.14~~	~~6.57~~	~~6.71~~	~~7.84~~	~~9.23~~	
S-1,2,4,5-TN	11.50	3.00	1.87	1.41	1.12	1.20	1.43	1.73	2.26	2.66	~~5.05~~	~~6.48~~	~~6.62~~	~~7.75~~	~~9.14~~	
S-1,2,3,5-TN	11.45	3.05	1.92	1.46	1.17	1.15	1.38	1.68	2.21	2.61	~~5.00~~	~~6.43~~	~~6.57~~	~~7.70~~	~~9.09~~	
S-1,2,4,6-TN	11.42	3.08	1.95	1.49	1.20	1.12	1.35	1.65	2.18	2.58	~~4.97~~	~~6.40~~	~~6.54~~	~~7.67~~	~~9.06~~	
S-2,3,4,5-TN	11.40	3.10	1.97	1.51	1.22	1.10	1.33	1.63	2.16	2.56	~~4.95~~	~~6.38~~	~~6.52~~	~~7.65~~	~~9.04~~	
S-1,2,3,4-TN	11.36	3.14	2.01	1.55	1.26	1.06	1.29	1.59	2.12	2.52	~~4.91~~	~~6.34~~	~~6.48~~	~~7.61~~	~~9.00~~	

Table 4.16 (Continued): Predicted retention times for sorbitol nitrates using the ensemble of 6xMLP:32-2-1

Compound	t_R^M (min) t_R^P (min)	14.50 Δt_R	13.37 Δt_R	12.91 Δt_R	12.62 Δt_R	10.30 Δt_R	10.07 Δt_R	9.77 Δt_R	9.24 Δt_R	8.84 Δt_R	6.45 Δt_R	5.02 Δt_R	4.88 Δt_R	3.75 Δt_R	2.36 Δt_R	
S-1,2,5,6-TN	11.23	3.27	2.14	1.68	1.39	0.93	1.16	1.46	1.99	2.39	~~4.78~~	~~6.21~~	~~6.35~~	~~7.48~~	~~8.87~~	Key*
S-2,3,6-TriN	7.93	~~6.57~~	~~5.44~~	~~4.98~~	~~4.69~~	2.37	2.14	1.84	1.31	0.91	1.48	2.91	3.05	~~4.18~~	~~5.57~~	25 % PI
S-2,3,5-TriN	7.89	~~6.61~~	~~5.48~~	~~5.02~~	~~4.73~~	2.41	2.18	1.88	1.35	0.95	1.44	2.87	3.01	~~4.14~~	~~5.53~~	50 % PI
S-4,5,6-TriN	7.84	~~6.66~~	~~5.53~~	~~5.07~~	~~4.78~~	2.46	2.23	1.93	1.40	1.00	1.39	2.82	2.96	~~4.09~~	~~5.48~~	75 % PI
S-3,4,6-TriN	7.83	~~6.67~~	~~5.54~~	~~5.08~~	~~4.79~~	2.47	2.24	1.94	1.41	1.01	1.38	2.81	2.95	~~4.08~~	~~5.47~~	80 % PI
S-1,3,4-TriN	7.77	~~6.73~~	~~5.60~~	~~5.14~~	~~4.85~~	2.53	2.30	2.00	1.47	1.07	1.32	2.75	2.89	~~4.02~~	~~5.41~~	90 % PI
S-1,2,3-TriN	7.76	~~6.74~~	~~5.61~~	~~5.15~~	~~4.86~~	2.54	2.31	2.01	1.48	1.08	1.31	2.74	2.88	~~4.01~~	~~5.40~~	95 % PI
S-3,4,5-TriN	7.74	~~6.76~~	~~5.63~~	~~5.17~~	~~4.88~~	2.56	2.33	2.03	1.50	1.10	1.29	2.72	2.86	~~3.99~~	~~5.38~~	99 % PI
S-2,5,6-TriN	7.67	~~6.83~~	~~5.70~~	~~5.24~~	~~4.95~~	2.63	2.40	2.10	1.57	1.17	1.22	2.65	2.79	~~3.92~~	~~5.31~~	
S-1,2,5-TriN	7.52	~~6.98~~	~~5.85~~	~~5.39~~	~~5.10~~	2.78	2.55	2.25	1.72	1.32	1.07	2.50	2.64	~~3.77~~	~~5.16~~	
S-1,2,6-TriN	7.51	~~6.99~~	~~5.86~~	~~5.40~~	~~5.11~~	2.79	2.56	2.26	1.73	1.33	1.06	2.49	2.63	~~3.76~~	~~5.15~~	
S-1,3,6-TriN	7.48	~~7.02~~	~~5.89~~	~~5.43~~	~~5.14~~	2.82	2.59	2.29	1.76	1.36	1.03	2.46	2.60	~~3.73~~	~~5.12~~	
S-3,5,6-TriN	7.43	~~7.07~~	~~5.94~~	~~5.48~~	~~5.19~~	2.87	2.64	2.34	1.81	1.41	0.98	2.41	2.55	3.68	~~5.07~~	
S-1,2,4-TriN	7.42	~~7.08~~	~~5.95~~	~~5.49~~	~~5.20~~	2.88	2.65	2.35	1.82	1.42	0.97	2.40	2.54	3.67	~~5.06~~	
S-2,3,4-TriN	7.39	~~7.11~~	~~5.98~~	~~5.52~~	~~5.23~~	2.91	2.68	2.38	1.85	1.45	0.94	2.37	2.51	3.64	~~5.03~~	
S-2,4,5-TriN	7.36	~~7.14~~	~~6.01~~	~~5.55~~	~~5.26~~	2.94	2.71	2.41	1.88	1.48	0.91	2.34	2.48	3.61	~~5.00~~	
S-1,3,5-TriN	7.35	~~7.15~~	~~6.02~~	~~5.56~~	~~5.27~~	2.95	2.72	2.42	1.89	1.49	0.90	2.33	2.47	3.60	~~4.99~~	
S-2,4,6-TriN	7.21	~~7.29~~	~~6.16~~	~~5.70~~	~~5.41~~	3.09	2.86	2.56	2.03	1.63	0.76	2.19	2.33	3.46	~~4.85~~	
S-2,3-DN	4.69	~~9.81~~	~~8.68~~	~~8.22~~	~~7.93~~	~~5.61~~	~~5.38~~	~~5.08~~	~~4.55~~	~~4.15~~	1.76	0.33	0.19	0.94	2.33	
S-2,5-DN	4.65	~~9.85~~	~~8.72~~	~~8.26~~	~~7.97~~	~~5.65~~	~~5.42~~	~~5.12~~	~~4.59~~	~~4.19~~	1.80	0.37	0.23	0.90	2.29	
S-3,4-DN	4.61	~~9.89~~	~~8.76~~	~~8.30~~	~~8.01~~	~~5.69~~	~~5.46~~	~~5.16~~	~~4.63~~	~~4.23~~	1.84	0.41	0.27	0.86	2.25	
S-1,3-DN	4.61	~~9.89~~	~~8.76~~	~~8.30~~	~~8.01~~	~~5.69~~	~~5.46~~	~~5.16~~	~~4.63~~	~~4.23~~	1.84	0.41	0.27	0.86	2.25	

Table 4.16 (Continued): Predicted retention times for sorbitol nitrates using the ensemble of 6xMLP:32-2-1

Compound	t_R^M (min) t_R^P (min)	14.50 Δt_R	13.37 Δt_R	12.91 Δt_R	12.62 Δt_R	10.30 Δt_R	10.07 Δt_R	9.77 Δt_R	9.24 Δt_R	8.84 Δt_R	6.45 Δt_R	5.02 Δt_R	4.88 Δt_R	3.75 Δt_R	2.36 Δt_R	
S-3,6-DN	4.53	~~9.97~~	~~8.84~~	~~8.38~~	~~8.09~~	~~5.77~~	~~5.54~~	~~5.24~~	~~4.71~~	~~4.31~~	1.92	0.49	0.35	0.78	2.17	Key*
S-3,5-DN	4.52	~~9.99~~	~~8.86~~	~~8.40~~	~~8.11~~	~~5.79~~	~~5.56~~	~~5.26~~	~~4.73~~	~~4.33~~	1.94	0.51	0.37	0.77	2.16	25 % PI
S-4,6-DN	4.45	~~10.05~~	~~8.92~~	~~8.46~~	~~8.17~~	~~5.85~~	~~5.62~~	~~5.32~~	~~4.79~~	~~4.39~~	2.00	0.57	0.43	0.70	2.09	50 % PI
S-1,4-DN	4.44	~~10.07~~	~~8.94~~	~~8.48~~	~~8.19~~	~~5.87~~	~~5.64~~	~~5.34~~	~~4.81~~	~~4.41~~	2.02	0.59	0.45	0.69	2.08	75 % PI
S-2,6-DN	4.41	~~10.09~~	~~8.96~~	~~8.50~~	~~8.21~~	~~5.89~~	~~5.66~~	~~5.36~~	~~4.83~~	~~4.43~~	2.04	0.61	0.47	0.66	2.05	80 % PI
S-4,5-DN	4.41	~~10.09~~	~~8.96~~	~~8.50~~	~~8.21~~	~~5.89~~	~~5.66~~	~~5.36~~	~~4.83~~	~~4.43~~	2.04	0.61	0.47	0.66	2.05	90 % PI
S-1,6-DN	4.39	~~10.11~~	~~8.98~~	~~8.52~~	~~8.23~~	~~5.91~~	~~5.68~~	~~5.38~~	~~4.85~~	~~4.45~~	2.06	0.63	0.49	0.64	2.03	95 % PI
S-1,2-DN	4.37	~~10.13~~	~~9.00~~	~~8.54~~	~~8.25~~	~~5.93~~	~~5.70~~	~~5.40~~	~~4.87~~	~~4.47~~	2.08	0.65	0.51	0.62	2.01	99 % PI
S-1,5-DN	4.34	~~10.16~~	~~9.03~~	~~8.57~~	~~8.28~~	~~5.96~~	~~5.73~~	~~5.43~~	~~4.90~~	~~4.50~~	2.11	0.68	0.54	0.59	1.98	
S-5,6-DN	4.32	~~10.18~~	~~9.05~~	~~8.59~~	~~8.30~~	~~5.98~~	~~5.75~~	~~5.45~~	~~4.92~~	~~4.52~~	2.13	0.70	0.56	0.57	1.96	
S-2,4-DN	4.27	~~10.24~~	~~9.11~~	~~8.65~~	~~8.36~~	~~6.04~~	~~5.81~~	~~5.51~~	~~4.98~~	~~4.58~~	2.19	0.76	0.62	0.52	1.91	
S-2-MN	2.08	~~12.43~~	~~11.30~~	~~10.84~~	~~10.55~~	~~8.23~~	~~8.00~~	~~7.70~~	~~7.17~~	~~6.77~~	~~4.38~~	2.95	2.81	1.68	0.29	
S-4-MN	2.04	~~12.46~~	~~11.33~~	~~10.87~~	~~10.58~~	~~8.26~~	~~8.03~~	~~7.73~~	~~7.20~~	~~6.80~~	~~4.41~~	2.98	2.84	1.71	0.32	
S-5-MN	2.02	~~12.48~~	~~11.35~~	~~10.89~~	~~10.60~~	~~8.28~~	~~8.05~~	~~7.75~~	~~7.22~~	~~6.82~~	~~4.43~~	3.00	2.86	1.73	0.34	
S-3-MN	2.01	~~12.49~~	~~11.36~~	~~10.90~~	~~10.61~~	~~8.29~~	~~8.06~~	~~7.76~~	~~7.23~~	~~6.83~~	~~4.44~~	3.01	2.87	1.74	0.35	
S-1-MN	1.91	~~12.59~~	~~11.46~~	~~11.00~~	~~10.71~~	~~8.39~~	~~8.16~~	~~7.86~~	~~7.33~~	~~6.93~~	~~4.54~~	3.11	2.97	1.84	0.45	
S-6-MN	1.88	~~12.62~~	~~11.49~~	~~11.03~~	~~10.74~~	~~8.42~~	~~8.19~~	~~7.89~~	~~7.36~~	~~6.96~~	~~4.57~~	3.14	3.00	1.87	0.48	

*Δt_R values highlighted in white – dark green depending on the PI, with darker green indicating values closer to measured t_R. Values outside of the 99 % PI shown in red and struck through

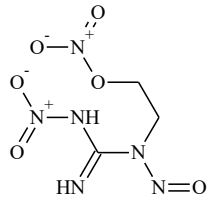
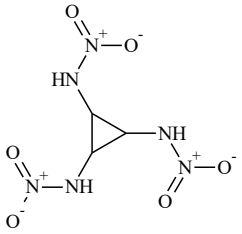
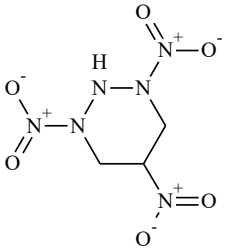
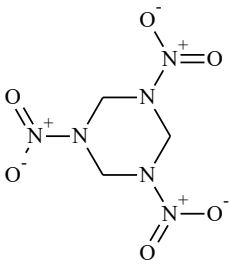
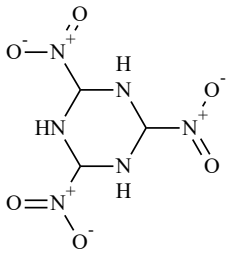
Table 4.17: Ions detected for each chromatographic peak in SHN sample

t_R^M (min)	Ion	Exact *m/z*	t_R^M (min)	Ion	Exact *m/z*
14.50	$C_6H_8O_{21}N_7^-$	513.9779	9.24	$C_6H_9O_{14}N_4^-$	361.0121
	$C_6H_8O_{16}N_5^-$	405.9972		$C_6H_{10}O_{14}N_4Cl^-$	396.9888
	$C_6H_9O_{14}N_4^-$	361.0121		$C_6H_{10}O_{12}N_3^-$	316.0270
13.37	$C_6H_9O_{16}N_5Cl^-$	441.9739	8.84	$C_6H_{10}O_{14}N_4Cl^-$	396.9888
	$C_6H_9O_{14}N_4^-$	361.0121		$C_6H_{10}O_{12}N_3^-$	316.0270
12.91	$C_6H_8O_{16}N_5^-$	405.9972	6.45	$C_6H_{10}O_{12}N_3^-$	316.0270
	$C_6H_9O_{16}N_5Cl^-$	441.9739		$C_6H_{11}O_{12}N_3Cl^-$	352.0037
	$C_6H_9O_{14}N_4^-$	361.0121		$C_6H_{11}O_{10}N_2^-$	271.0419
12.62	$C_6H_9O_{16}N_5Cl^-$	441.9739		$C_6H_{12}O_{10}N_2Cl^-$	307.0186
	$C_6H_9O_{14}N_4^-$	361.0121	5.02	$C_6H_{11}O_{12}N_3Cl^-$	352.0037
10.30	$C_6H_{10}O_{14}N_4Cl^-$	396.9888		$C_6H_{11}O_{10}N_2^-$	271.0419
	$C_6H_{10}O_{12}N_3^-$	316.0270	4.88	$C_6H_{10}O_{12}N_3^-$	316.0270
10.07	$C_6H_9O_{14}N_4^-$	361.0121		$C_6H_{11}O_{10}N_2^-$	271.0419
	$C_6H_{10}O_{14}N_4Cl^-$	396.9888	3.75	$C_6H_{10}O_{12}N_3^-$	316.0270
	$C_6H_{10}O_{12}N_3^-$	316.0270		$C_6H_{11}O_{12}N_3Cl^-$	352.0037
	$C_6H_{11}O_{12}N_3Cl^-$	352.0037		$C_6H_{11}O_{10}N_2^-$	271.0419
9.77	$C_6H_9O_{14}N_4^-$	361.0121	2.36	$C_6H_{11}O_{10}N_2^-$	271.0419
	$C_6H_{10}O_{14}N_4Cl^-$	396.9888		$C_6H_{12}O_{10}N_2Cl^-$	307.0186
	$C_6H_{10}O_{12}N_3^-$	316.0270			

4.3.5.4 Excluding compounds with the same elemental composition

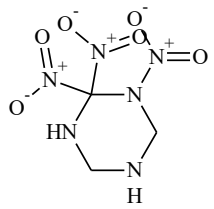
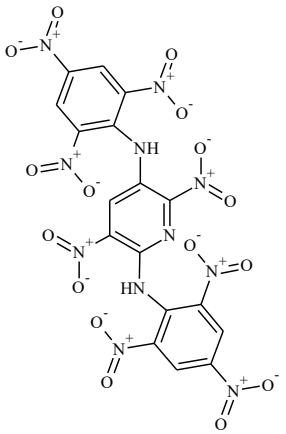
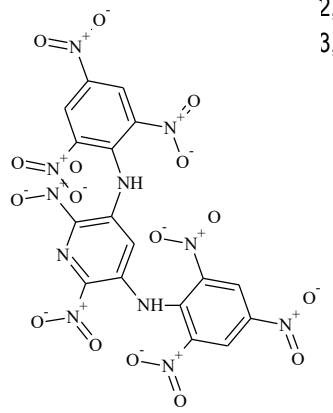
In addition to predicting the retention times of explosives, retention time prediction could be of value for excluding other compounds that share the same elemental composition. To test this theory, the ensemble (6xMLP:32-2-1) was used to predict the retention times for all compounds found in Chemspider with the same elemental composition as three of the explosives, RDX, PYX and TATB, included in the test set for this model. There were six compounds (including RDX) with an elemental composition of $C_3H_6O_6N_6$, three compounds (including PYX) with an elemental composition of $C_{17}H_7O_{16}N_{11}$ and nine compounds (including TATB) with an elemental composition of $C_6H_6O_6N_6$ in the Chemspider database (Table 4.18).

Table 4.18: Predicted retention times for compounds with the same elemental composition as RDX ($C_3H_6O_6N_6$), PYX ($C_3H_6O_6N_6$) and TATB ($C_6H_6O_6N_6$).

Elemental composition	Chemical structure	Systematic name	6xMLP:32-2-1 t_R^P (min)	6xMLP:32-2-1 Δt_R (min)
$C_3H_6O_6N_6$		2-(N'-Nitro-N-nitrosocarbamimidamido)ethyl nitrate	2.64	-3.31
$C_3H_6O_6N_6$		N,N',N''-Trinitro-1,2,3-cyclopropanetriamine	2.64	-3.31
$C_3H_6O_6N_6$		1,3,5-Trinitro-1,2,3-triazinane	3.90	-2.05
$C_3H_6O_6N_6$		1,3,5-Trinitro-1,3,5-triazinane[a]	4.04[a]	-1.91[a]
$C_3H_6O_6N_6$		2,4,6-Trinitro-1,3,5-triazinane	4.15	-1.80

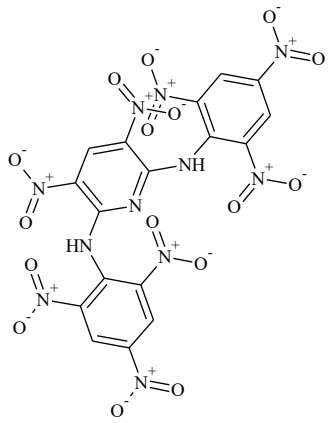
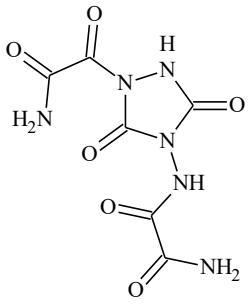
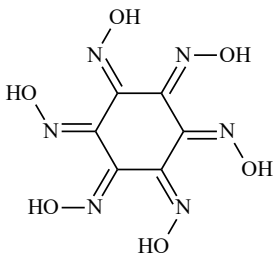
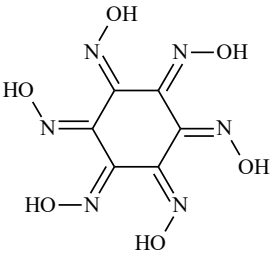
[a]RDX; [b]PYX; [c]TATB

Table 4.18 (Continued): Predicted retention times for compounds with the same elemental composition as RDX ($C_3H_6O_6N_6$), PYX ($C_3H_6O_6N_6$) and TATB ($C_6H_6O_6N_6$).

Elemental composition	Chemical structure	Systematic name	6xMLP:32-2-1 t_R^P (min)	6xMLP:32-2-1 Δt_R (min)
$C_3H_6O_6N_6$		1,2,2-Trinitro-1,3,5-triazinane	5.23	-0.72
$C_{17}H_7O_{16}N_{11}$		3,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-2,5-pyridinediamine	17. 98	0.99
$C_{17}H_7O_{16}N_{11}$		2,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-3,5-pyridinediamine	17.95	1.05

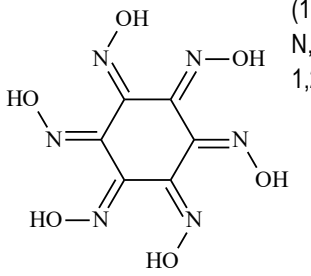
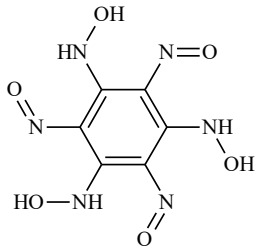
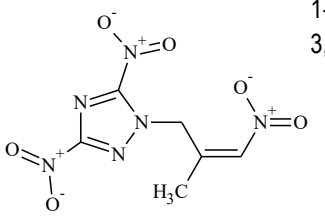
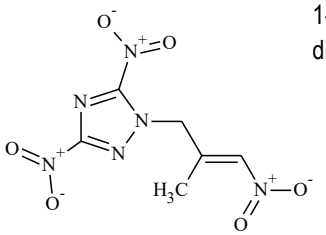
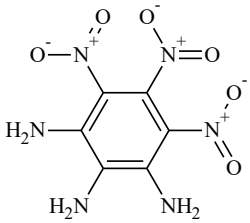
[a]RDX; [b]PYX; [c]TATB

Table 4.18 (Continued): Predicted retention times for compounds with the same elemental composition as RDX ($C_3H_6O_6N_6$), PYX ($C_3H_6O_6N_6$) and TATB ($C_6H_6O_6N_6$).

Elemental composition	Chemical structure	Systematic name	6xMLP:32-2-1 t_R^P (min)	6xMLP:32-2-1 Δt_R (min)
$C_{17}H_7O_{16}N_{11}$		3,5-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-,6-pyridinediamine[b]	18.08[b]	1.18[b]
$C_6H_6O_6N_6$		N-{1-[Amino(oxo)acetyl]-3,5-dioxo-1,2,4-triazolidin-4-yl}ethanediamide	0.13	-9.27
$C_6H_6O_6N_6$		(1Z,2Z,3E,4Z,5Z,6E)-N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	3.30	-6.10
$C_6H_6O_6N_6$		N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	3.30	-6.10

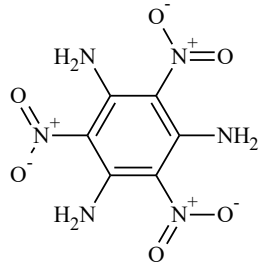
[a]RDX; [b]PYX; [c]TATB

Table 4.18 (Continued): Predicted retention times for compounds with the same elemental composition as RDX ($C_3H_6O_6N_6$), PYX ($C_3H_6O_6N_6$) and TATB ($C_6H_6O_6N_6$).

Elemental composition	Chemical structure	Systematic name	6xMLP:32-2-1 t_R^P (min)	6xMLP:32-2-1 Δt_R (min)
$C_6H_6O_6N_6$		(1E,2E,3E,4E,5E,6E)-N,N',N'',N''',N'''',N''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	3.30	-6.10
$C_6H_6O_6N_6$		N,N',N''-Trihydroxy-2,4,6-trinitroso-1,3,5-benzenetriamine	3.39	-6.02
$C_6H_6O_6N_6$		1-[(2Z)-2-Methyl-3-nitro-2-propen-1-yl]-3,5-dinitro-1H-1,2,4-triazole	8.75	-0.65
$C_6H_6O_6N_6$		1-(2-Methyl-3-nitro-2-propen-1-yl)-3,5-dinitro-1H-1,2,4-triazole	8.93	-0.47
$C_6H_6O_6N_6$		4,5,6-Trinitro-1,2,3-benzenetriamine	11.55	2.15

[a]RDX; [b]PYX; [c]TATB

Table 4.18 (Continued): Predicted retention times for compounds with the same elemental composition as RDX ($C_3H_6O_6N_6$), PYX ($C_3H_6O_6N_6$) and TATB ($C_6H_6O_6N_6$).

Elemental composition	Chemical structure	Systematic name	6xMLP:32-2-1 t_R^P (min)	6xMLP:32-2-1 Δt_R (min)
$C_6H_6O_6N_6$		2,4,6-Trinitro-1,3,5-benzenetriamine[c]	11.92[c]	2.52[c]

[a]RDX; [b]PYX; [c]TATB

As shown in Figure 4.14, using the ensemble of 6xMLP:32-2-1, all six compounds with an elemental composition of $C_3H_6O_6N_6$ had a predicted retention time within the 99 % PI of the measured retention time of RDX. Only one of the six compounds had a predicted retention time within the 75 % PI, but this was not RDX. Therefore, in this instance the model was not able to reduce the number of possible compounds. Similarly, this model was not able to distinguish between the three compounds with an elemental composition of $C_{17}H_7O_{16}N_{11}$ as all three compounds had a predicted retention time within the 75 % PI of the measured retention time for PYX. This was perhaps unsurprising given the structural similarity of these three compounds, which only differed in the location of the amine and nitro groups on the central pyridine ring (Table 4.18). Five of the nine compounds with an elemental composition of $C_6H_6O_6N_6$ did have a

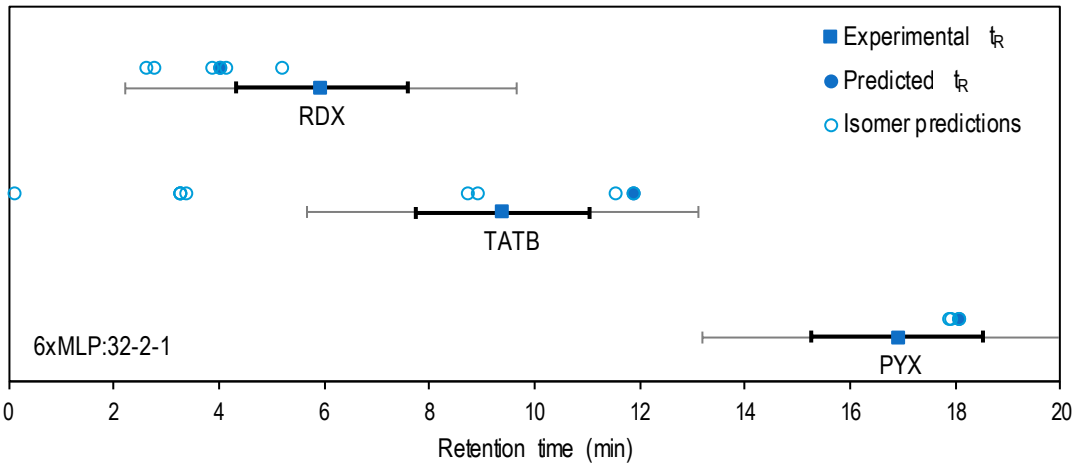


Figure 4.14: Experimental and predicted retention times for RDX, TATB and PYX compared to predicted retention times of structural isomers using A) ensemble of 6xMLP:32-2-1 and B) MLP:11-3-1. Grey error bars show 99 % PI and black error bars show 75 % PI.

predicted retention time outside of the 99 % PI from the measured retention time obtained for TATB and so could potentially be excluded. There remains approximately a 1 % probability that the retention model would lead to prediction errors outside of the 99 % PI. However, given that the predicted retention times for these five compounds were ≥ 2.5 min less than the 99 % PI, it was thought unlikely that any of these compounds would have the same retention time as TATB. Two compounds fell within the 75 % PI, but neither were TATB. For confirmatory identification of TATB, the chromatographic retention time and/or MS/MS spectra need to differ from the other 8 isomers. Therefore, it may be necessary to show that the other three compounds that fell within the 99 % PI, had different measured retention times and/or MS/MS spectra, regardless of whether or not they fell within the 75 % PI.

4.3.6 Prediction of retention times in matrix

For trace analysis of explosives, samples will likely be received in complex matrices. Munro et al. highlighted the importance of using matrix-matched standards for reliable ANN modelling of pharmaceutical retention times in wastewater [83]. Here the effect of two matrices, latent human fingermarks and pond water, on the measured retention times of 149 explosives and drugs was investigated. One explosive (HNDPA) and thirteen drugs were not detected in the spiked pond water sample and nine drugs were not detected in the spiked fingermark samples, presumably due to matrix suppression. In the fingermark samples, the average (n=6) retention times for 94 % of detected explosives (44 out of 47) and 98 % of detected drugs (91 out of 93) were within 2.5 % of the average retention time in solvent. For the pond water samples, the average (n=3) retention times for 98 % of detected explosives (45 out of the 46) and 52 % of detected drugs (46 out of 89) were within 2.5 % of the average retention time in solvent. The effect of the two matrices on retention time is shown in Figure 4.15, with x-error bars indicating standard deviation in solvent and y-error bars indicating standard deviation in matrix. In the majority of cases, the pond water matrix had a greater effect on retention time than the fingermark matrix. Notable exceptions to this were three explosives PA, PYX and HDNPA (not detected in pond water) which had the least reproducible retention times in solvent (standard deviation (SD) = 0.11, 0.22 and 0.13 min respectively). The biggest differences in retention time were seen for HMDD (Δt_R = -1.07 min in spiked pond water, t_R = 4.43 min in solvent), bumetanide (Δt_R = +1.15 min in spiked pond water, t_R = 6.25 min in solvent) and piretanide (Δt_R = +1.18 min in spiked pond water, t_R = 5.12 min in

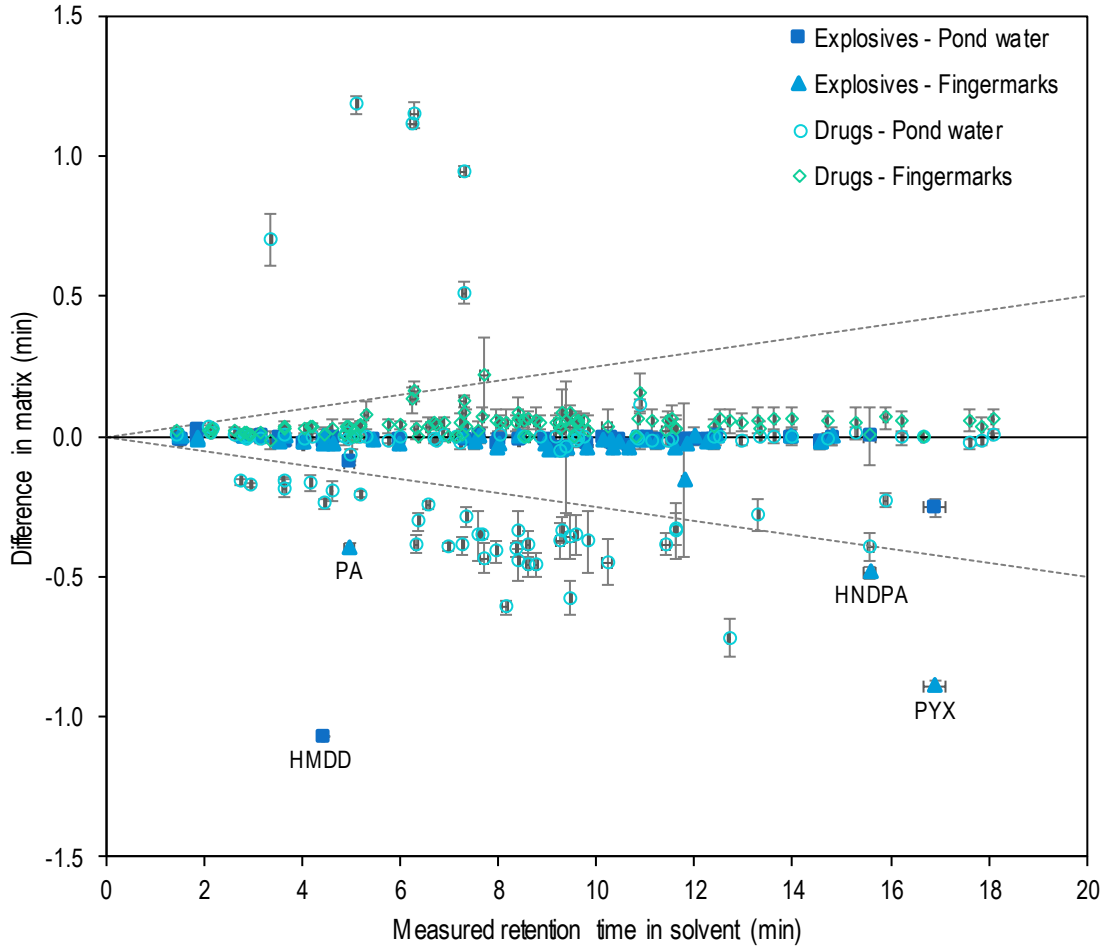


Figure 4.15: Effect of pond water and fingermark matrices on retention times for drugs and explosives. Vertical error bars indicate standard deviation in matrix (n=3 for pond water, n=6 for fingermarks) and x-error bars indicate standard deviation in solvent (n=6). Dashed grey lines show ± 2.5 % retention time in solvent.

solvent). Ideally matrix matched standards should be used to train the ANN model. Due to the vast range of matrices encountered during trace-explosive forensic casework, this may pose a challenge. Without matrix-matched standards, or suitable sample clean-up to remove matrix effects, larger prediction errors would be expected for some compounds due to the effect of matrix on measured retention time and larger prediction intervals would be required to account for uncertainties in this matrix effect.

4.4 Conclusion

Average retention times of 149 compounds (47 explosives and 102 drugs) were obtained for a gradient LC separation on a C_{18} Ar column. ANN models were initially developed for this dataset using three sets of descriptors previously selected for QSRR models on C_{18} columns. Models using descriptors selected by Aalizadeh et al. and Barron & McEneff performed better than the best model using descriptors selected by Goryński et al. However, for both of these models, overall prediction errors were larger for this dataset, than for the best models presented in the original publications for the prediction of retention time on a C_{18} column, perhaps due to additional retention mechanisms on a C_{18} Ar column. MLR models were also investigated using the descriptors selected by Goryński et al. and Aalizadeh et al. but performance was poorer than for the ANN models developed using these same descriptors. Two sets of descriptors were selected specifically for this dataset and LC separation. The first contained 32 descriptors which were a subset from the descriptors selected by forward selection and the second contained 11 descriptors which were a subset from the descriptors initially selected by genetic algorithm. In both cases, a 3-layer MLP was the best performing neural network. Prediction errors for the best models developed using these descriptors were lower than prediction errors for the models using previously selected descriptors. An ensemble of six repeated networks was formed using both sets of descriptors and no significant difference was found between the ensemble and best MLP:32-2-1 network, but prediction errors were significantly larger for the ensemble than best MLP:11-3-1 network. Predicted retention order was investigated for the 6xMLP:32-2-1 ensemble and individual MLP:11-3-1 network. Significant correlation between predicted retention order and measured retention order was found, but even when just looking at the nitrate esters, there were cases with both models where the predicted retention order was incorrect. While predicted retention times may provide an initial indication, it would be necessary to confirm predicted retention orders with reference standards. Prediction intervals (PI) were calculated to indicate confidence in predicted retention times and the 75 % and 99 % PI were approximately ± 1.6 min and ± 3.7 mins, respectively for both MLP:11-3-1 and 6xMLP:11-3-1. Based on the MAE, RMSE and PIs, determined for the original 149 compounds used for training, selection and testing; performance of the 6xMLP:32-2-1 ensemble and MLP:11-3-1 network was comparable, and it was not possible to choose one model over the other. However, when used to predict retention times for six new OGSR compounds, the 6xMLP:32-2-1 was clearly superior with the MAE and

RMSE for 6xMLP:32-2-1 less than half those for MLP:11-3-1. Additionally, half of the retention times predicted using MLP:11-3-1 fell outside of the corresponding 99 % PI; whereas, the predicted retention times for all six new compounds fell within the 95 % PI when 6xMLP:32-2-1 was used. Therefore, 6xMLP:32-2-1 was selected for prediction of retention times for the remaining new compounds, due to better generalisability of the model and associated PIs. The predicted retention times for all MEKP oligomers also fell within the 99 % PIs. Retention time prediction for mannitol and sorbitol nitrates, and the use of 99 % PI for exclusion purposes, supported identification of sugar nitrates where the $[M-H]^-$ ion of a less-nitrated sugar and the $[M-NO_2]^-$ ion of a more-nitrated sugar would have the same exact mass. Additionally, 99 % PI were used to exclude some TATB isomers. However, measured retention times would still be required to confirm identification and in the case of RDX and PYX, retention time prediction did not reduce the number of possible compounds with the same elemental composition as all predicted retention times fell within the 99 % PI. For more confident exclusion of non-explosives, further testing using a wider range of non-explosive compounds, than the drugs used here, and a larger number of blind test compounds may be required to further assess the generalisability of the PIs used here. The majority of explosives retention times were unaffected by either the fingermark or pond water matrices. However, the potential effect of matrix on the retention times for both explosive and non-explosive compounds should be considered if attempting to use a prediction model for exclusion purposes. Overall, the prediction models developed here may support the preliminary identification of suspect or unknown explosives detected by LC-HRMS and provide the potential to eliminate some compounds with the same elemental composition, reducing the time and cost implications of purchasing/synthesising and analysing reference materials for elimination purposes.

Chapter 5: Conclusions and future work

5.1 Summary

The aim of this thesis was to investigate the suitability of LC-HRMS for screening and identification of energetic materials. Chapter 1, examined the status quo of screening and identification of energetic materials, for which targeted LC-MS methods have often been used. The requirement for forensic scientists to be able to detect and identify any explosive substance was also discussed, along with the potential value of LC-HRMS for targeted analysis, suspect screening and non-target analysis to assist forensic scientists in meeting this requirement. Additional benefits, such as the potential to include more compounds in one method and achieve greater confidence of identification with LC-HRMS, were also discussed. Several challenges to utilising the potential of LC-HRMS were recognised. These included the challenges of: (a) achieving chromatographic separation and simultaneous ionisation, in order to include more compounds in one method; (b) setting evidence-based identification criteria and assessing the greater confidence of identification provided by LC-HRMS; (c) optimising and performance testing LC-HRMS method for suspect and non-target explosive substances and (d) how suspect or non-target explosive substances can be identified without reference materials, at least in the first instance.

The first challenges, achieving chromatographic separation and simultaneous ionisation, were addressed in Chapter 2, where the objective was to develop a liquid chromatography – high resolution mass spectrometry (LC-HRMS) method for screening multiple classes of explosives. Sixteen nitro-explosives, including nitrate esters, nitroaromatics and nitramines, plus two peroxides, were selected for method development and optimisation. Chromatographic separation for the existing method, on a PFP column, was improved by using a temperature gradient, but the best separation was achieved through gradient elution on a C_{18}Ar column, at a fixed temperature of 20 °C. With ESI and the ammonium acetate mobile phase additive, used for a previous LC-MS/MS method, here no acetate adduct was detected for nitroglycerin, despite it being detected with the previous method. This demonstrated the need to optimise MS conditions for specific instruments, rather than simply transferring methods, especially when detecting adduct ions where stability may be affected by differences in instrumentation. Additionally, nitrobenzene and the nitrotoluenes were not detected at all with ESI. Ionisation of all 18 initial target analytes was achieved using ammonium chloride and APCI. Negative mode APCI was used for the nitro explosives and chloride adducts were detected for the nitramines and nitrate esters. The

peroxides were detected in positive mode APCI and ammonium adducts formed for TATP. However, compromise ionisation conditions, particularly in terms of capillary and vaporiser temperatures, were required to allow detection of all the target analytes. Greater sensitivity could be achieved using conditions optimised for individual classes of explosive. For example, the nitrate esters favoured lower temperatures than those selected here and were detected at lower concentrations using ESI (pg versus ng levels).

The challenge of setting evidence-based identification criteria and assessing the greater confidence of identification provided by LC-HRMS was addressed in Chapter 3. While HRMS offers greater confidence of identification due to the ability to separate isobaric ions, which are indistinguishable by low-resolution MS, a lack of consensus and scientific evidence to support the selection of identification criteria for LC-HRMS analysis of energetic materials was identified. Theoretical experiments were performed to determine the number of theoretical elemental compositions within 1, 2 and 5 ppm of the monoisotopic mass of 60 explosive substances and the number of isomers with the same elemental composition. Mass accuracy had a greater effect on the number of theoretical elemental compositions of larger compounds with 223 theoretically possible elemental compositions within 5 ppm of PYX. Even with a mass accuracy threshold of 1 ppm, 23 compounds had more than one theoretically possible elemental composition. While narrower mass accuracy thresholds and isotope profiles could reduce the number of possible elemental compositions, they could not provide any discrimination between isomers. The number of other isomers listed in the ChemSpider database ranged from none for PETN and TEGDN to 10,083 for DEDPU. This illustrated the need to consider the compounds being identified when setting identification requirements, since confirming the presence of DEDPU rather than one of the other 10,083 isomers would be much more challenging than confirming the presence of PETN or TEGDN. Regardless, detection of a single HRMS ion is unlikely to be sufficient for identification. Using APCI, a range of different ions were detected, including in-source fragment ions and adduct ions, in addition to molecular ions and (de)protonated molecules. Therefore, the elemental composition of a detected ion could be due to a fragment of a larger molecule as well as isomers with the same elemental composition. The effect of identification requirements, such as the number and type of ions and retention time windows, on selectivity (both between explosives and from fingermark, indoor dirt and outdoor dirt matrices) and detection limits for explosives was also investigated. Several cases of different explosive compounds producing common ions were

identified, with 15 compounds producing a nitrate ion and HMX producing ions with the same *m/z* as all three RDX ions. In some cases, the use of multiple ions led to improved selectivity between different explosive compounds and from matrix. However, with vastly poorer detection limits for less abundant ions, in some cases, this came at the cost of reduced sensitivity. The greatest improvement in selectivity was seen when a retention time window was applied. Even with a 2.5 % retention time window, there were a few cases of selectivity issues between pairs of isomers, but in the case of 2,4- and 2,6-DNT the use of ion ratios provided sufficient discrimination. Due to the compromise between selectivity and sensitivity, recommended identification requirements would vary depending on how LC-HRMS was used. For use as a screening method, where positive results would be confirmed using another method and minimising the number of false negatives was the priority, accurate mass of the most abundant ion and retention time may be sufficient. For confirmatory analysis, even with LC-HRMS, detection of precursor ion, multiple fragment ions and calculation of corresponding ion ratios may be required to provide sufficient discrimination between isomers and other compounds producing common ions.

The challenge of optimising and performance testing an LC-HRMS method for suspect and non-target explosive substances was addressed by optimising the method for a set of 18 initial target analytes in Chapter 2 and then assessing generalisability and performance for an additional 44 compounds in Chapter 3, for which no method optimisation had been performed, to mimic suspect or non-target compounds. Overall, the LC-HRMS method generalised well to the larger set of target compounds, with at least one ion detected at 10 ng μL^{-1} in all cases. Additionally, for 54 compounds, at least three ions (including the molecular ion/(de)protonated molecule/adduct ion and a fragment ion) were detected, even without the use of HCD fragmentation. However, signal to noise ratios and LODs varied greatly for different compounds (from low pg to ng levels) and for different ions of the same compound. Multiple oligomers of MEKP were also detected, as were mannitol and sorbitol penta- and tetra- nitrates and sorbitol tri- and di- nitrates. In Chapter 4, an additional eight organic gunshot residue compounds were also detected with baseline chromatographic separation between the two nitro- and four dinitro- diphenylamine isomers, without any further chromatographic optimisation required.

The final challenge, preliminary identification of suspect or non-target explosive substances was addressed in Chapter 4, where prediction of chromatographic retention times was investigated for the LC method optimised in Chapter 2. This was because prediction of retention times for

isomers with the same elemental composition may enable prioritisation and/or reduction of the number of compounds for which reference materials need to be purchased and/or synthesised. Average measured retention times for 149 compounds (47 explosives and 102 drugs) were obtained for the gradient LC separation on the C_{18} Ar column and automatically split into training, selection and test sets when used to develop prediction models. Initially, the suitability of three previously selected descriptor sets (used by Goryński et al., Barron & McEneff and Aalizadeh et al.) was investigated using machine learning (specifically artificial neural networks, ANN) for the prediction of explosive retention times for this LC method. The models using descriptors selected by Aalizadeh et al. and Barron & McEneff performed better than the best model using descriptors selected by Goryński et al. However, prediction errors were still greater for this dataset, than those presented in the original publications. This suggested that descriptors selected for the prediction of retention times on a C_{18} column, may not fully describe retention mechanisms on a C_{18} Ar column. Multiple linear regression (MLR) models were also investigated, but performance was poorer than for the equivalent ANN model. Two sets of molecular descriptors were then selected specifically for this LC method and set of analytes. The first set of descriptors were selected by forward selection and a subset of 32 descriptors produced the best prediction model (MLP:32-2-1) and ensemble (6xMLP:32-2-1). The second set of descriptors were selected by genetic algorithm and in this case a subset of 11 descriptors led to the best prediction model (MLP:11-3-1). Prediction errors for the best models using both these sets of descriptors were lower than the prediction errors for models using previously published descriptors. This suggested that it was worth selecting descriptors specifically for the LC method used, rather than simply using previously published descriptor sets. Prediction intervals (PI) were calculated to indicate confidence in predicted retention times and the 75 and 99 % PI were approximately ± 1.6 and ± 3.7 mins, respectively, for the best models. Based on the MAE, RMSE and PIs, determined for the original 149 compounds used for training, selection and testing, performance of the 6xMLP:32-2-1 and MLP:11-3-1 models was comparable, and it was not possible to choose one over the other. However, when an external test set was used the 6xMLP:32-2-1 model was clearly superior, with the MAE and RMSE for 6xMLP:32-2-1 less than half those for MLP:11-3-1. This demonstrated the importance of testing prediction models using an external test set, rather than simply using the internal test set which, although not used for training, was still considered during model selection. The 6xMLP:32-2-1 model demonstrated good generalisability to the external test set, with all predicted retention times falling within the 95 % PI. Therefore, 6xMLP:32-2-1 was

selected for prediction of retention times for the remaining new compounds. The predicted retention times for all MEKP oligomers fell within the 99 % PIs. Retention time prediction for mannitol and sorbitol nitrates, and the use of 99 % PI for exclusion purposes, supported identification of sugar nitrates where the $[M-H]^-$ ion of a less-nitrated sugar and the $[M-NO_2]^-$ ion of a more-nitrated sugar would have the same exact mass. Additionally, 99 % PI were used to exclude some TATB isomers, supporting the hypothesis that retention time prediction may enable prioritisation and/or reduction of the number of compounds for which reference materials need to be purchased and/or synthesised. However, in the case of RDX and PYX, retention time prediction did not reduce the number of possible compounds with the same elemental composition, as all predicted retention times fell within the 99 % PI. Therefore, retention time prediction may be of value in some, but not all, cases. For confirmatory identification, measured retention times and MS/MS would still be required, especially for forensic purposes.

Application of the LC-HRMS method to the detection of explosives in a number of matrices was also demonstrated. The method was used to detect explosives in contact traces and 19 nitro-explosives were detected in spiked fingermarks. RDX and HMX chloride adducts and RDX fragment ions were also detected in extracted ion chromatograms for a fingermark depletion series, following handling of a commercial explosive. The results from this depletion series also highlighted the subjectivity of low level peak detection in LC-HRMS, when there is often no background noise for determination of a signal-to-noise ratio. The method was also successfully used to screen for explosives on a passive vapour sampler, with 12 explosive compounds detected following a 24-hour exposure in an explosive magazine. For the DNBs, DNTs, RDX and TNT, multiple ions were detected and at least one ion ratio fell within the relevant tolerance from average ion ratios calculated in solution. Ion ratios could also be used to distinguish DNB and DNT isomers. Finally, retention times for the 47 explosives used to develop prediction models were also obtained in fingermark and pond water matrices. In the majority of cases retention times were not affected by matrix. The only exceptions were PA, HMDD, HNDPA and PYX for which matrix matched standards may be required.

5.2 Conclusions and recommendations

1. LC-HRMS can be used to include more explosives in one method and here the ability to separate chromatographically and simultaneously ionize multiple classes of explosives has been demonstrated. However, due to different ionization conditions resulting in the greatest sensitivity for different classes of explosives, this came at a cost of reduced sensitivity for some compounds.
2. Generalizability of an LC-HRMS method, developed and optimized for 18 target explosives, to a larger set of energetic materials has also been demonstrated for the first time. This supports the potential use of full-scan LC-HRMS, for suspect screening and non-target analysis of energetic materials.
3. Good chromatographic separation and tandem mass spectrometry are recommended for confirmatory analysis, along with a consideration of the number of isomers for individual compounds. While in-source fragment ions and the use of ion ratios were used to increase selectivity; greater confidence of identification might be achieved through isolation of a precursor ion for fragmentation, rather than simply relying on chromatographic retention time to link precursor and fragment ions. Here, it was not always possible to unequivocally identify energetic materials using LC-HRMS, due to the presence of hundreds or even thousands of isomers in some cases. Even using LC-MS/HRMS it may not be possible to eliminate all 10,083 other isomers and unequivocally identify DEDPU. Additionally, for some compounds, there were multiple possible elemental compositions within the achievable mass accuracy and resolution. This was further complicated by the fact that, with APCI, a measured accurate mass could have been due to a fragment and/or adduct ion, as well as a molecular ion and/or (de)protonated molecule.
4. While reference materials would be required for confirmatory identification of explosive substances, the ANNs developed in this thesis may be of value for prioritizing the acquisition of reference materials following detection of suspect or non-target energetic materials. Additionally, the use of 99 % prediction intervals for ANNs demonstrated potential for the exclusion of compounds with the same elemental composition, but less than a 1 % chance of having the same retention time.

5.3 Recommendations for future work

5.3.1 Validation of LC-HRMS method and ANN prediction model

While method performance has been assessed here, full method validation, to demonstrate that the method fulfils specified requirements and is adequate for its intended use, would be required before the LC-HRMS method could be used for routine forensic analysis. As discussed in Chapter 3, there are currently no agreed validation guidelines specifically for LC-HRMS analysis of energetic materials, and so before validation could be performed an agreement on the performance criteria that must be met would be needed. Validating LC-HRMS methods for suspect screening and non-targeted analysis presents further challenges as reference materials are not normally analysed in advance. One approach could be to obtain reference materials and validate the method for all suspect compounds. Whilst time consuming and costly, this would also present an opportunity to build a larger database of reference spectra and retention times obtained using the LC-HRMS method in question. The database could then be used to support the preliminary identification of suspect compounds, for which reference materials are not run during routine analysis. Of course, this would still not be possible for non-target compounds which are unknown to the analyst in advance.

If the LC-HRMS method was validated for additional suspect energetic materials, this would provide a larger number of external blind test compounds with measured retention times that could be used to further assess accuracy of the retention time prediction model developed in Chapter 4 for new energetic materials. Further validation of the prediction model and prediction intervals would also be required for more confident exclusion of non-explosives. The number of compounds used for ANN modelling here was also relatively small, especially in comparison to the number of known compounds (>63 million compounds in the ChemSpider database). Further testing using a wider range of non-explosive compounds, than the drugs used here, would be required to further assess the generalisability of the prediction intervals used here. Determination of the applicability domain of the model may also increase confidence in predicted results, through identification of compounds which are outliers.

5.3.2 Non-target screening

Here, the suitability of LC-HRMS has only been demonstrated for targeted and suspect screening, against a list of target or suspect explosives. To truly take advantage of the power of full-scan HRMS, it would be desirable to perform non-target analysis enabling the detection and

identification of unknown explosive substances. However, further work is required before this would be possible. Challenges remain regarding locating unknown peaks of interest amongst the large amount of data collected by full-scan HRMS. Since not all components of a sample are of interest in the forensic analysis of explosives, it is biased non-target screening [72], involving the discovery of new compounds related to known energetic materials that is of interest to forensic scientists. Two approaches have been used for discovery of relevant components in biased non-target screening of new psychoactive substances (NPS), a top-down and a bottom-up approach [106]. The top-down approach, where abundant peaks are selected from the total ion chromatogram, may not be successful for discovery of trace explosives, since the explosive component often has a low abundance in comparison to matrix components. The bottom-up approach, where extracted ion chromatograms are generated for common fragment ions or neutral losses, may offer greater promise for biased non-target screening of energetic materials and warrants further investigation. In particular, use of the common fragment ions identified in Chapter 3 could be investigated. Once peaks of interest have been discovered the next challenge is to identify the relevant compound. Use of APCI makes identification of unknowns more challenging than with ESI. This is because ions may be produced with odd or even numbers of electrons, since both $[M]^{-\cdot}$ and $[M-H]^-$ ions can form by APCI. Determination of a molecular formula is further complicated by the presence of adduct and in-source fragment ions. Therefore, development of prediction models to predict the type of ions an analyte would form could also be of value here.

5.3.3 Optimisation of sample preparation for suspect screening and non-target analysis

This thesis has predominantly focussed on the LC-HRMS method. However, having a ‘universal’ LC-HRMS method for suspect or non-target screening is only of value if the sample preparation method used is also ‘universal’. This is because only compounds that are recovered and extracted into solution can be detected by LC-HRMS. Sample preparation techniques, such as solid phase extraction (SPE), offer the potential for improved sensitivity due to a concentration factor. Additionally, suitable sample preparation techniques may be able to remove matrix, reducing/eliminating the effect of matrix on retention time and/or ion ratios and hence simplify identification. However, if sample preparation methods are too selective then some suspect or non-target compounds of interest may also be removed.

Research into sample preparation for suspect screening has already begun. With me as a co-author, Rapp-Wright et al. assessed recoveries of the initial 18 target explosives used in this thesis, from wastewater, using 34 different SPE sorbents [130]. Suitability of the optimised SPE method, using Oasis HLB (divinylbenzene with n-vinylpyrrolidone) cartridges, for suspect screening was assessed by determining recoveries of an additional 17 explosives-related analytes which were not considered during optimisation. Recoveries for 15 of the 17 additional analytes were greater than 50 %, but the remaining two compounds, NQ and PA were not recovered [130]. This demonstrated the potential for suspect screening using SPE followed by LC-HRMS. However, a truly 'universal' sample preparation and LC-HRMS screening method is still yet to be achieved and assessment of recoveries from a wider range of forensically relevant matrices would be required.

5.3.4 LC-MS/HRMS

For increased confidence of identification, LC-MS/HRMS could be investigated. In comparison to the in-source fragment ions used in this thesis, which could also result from coeluting compounds; LC-MS/HRMS increases the confidence of identification through linking high resolution fragment ions to a specific precursor *m/z*. Traditional, tandem mass spectrometry in multiple reaction monitoring (MRM) mode would not be suitable for suspect or non-target screening as only ions with preselected masses would be detected, but it could be useful for subsequent confirmatory analysis. For non-target screening there are two main approaches to the use of LC-MS/HRMS, data-independent and data-dependent acquisition [145]. With data-independent LC-MS/HRMS, typically all ions are fragmented, and so product and precursor ions can only be linked based on matching chromatographic retention times, as is the case with in-source fragmentation. To somewhat overcome this limitation, Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH) has been used in toxicology [146]. In this case the full *m/z* range is divided into fixed smaller *m/z* ranges which are each independently fragmented. Therefore, while product ions still cannot be directly linked to a specific precursor *m/z*, they can be linked to smaller *m/z* ranges. Data-dependent LC-MS/HRMS, where acquisition switches from HRMS to MS/HRMS only when ions are detected above a threshold abundance, has also been used in toxicology [147]. In this case, product ions can be directly linked to a specific precursor *m/z*, but only precursor ions detected above the threshold abundance will be fragmented. All in all, further work is required to investigate the additional value of LC-MS/HRMS compared to LC-HRMS and

the best modes of acquisition for targeted analysis, suspect screening and non-target analysis of energetic materials.

5.3.5 Comprehensive two-dimensional chromatography

In this thesis, a good chromatographic separation, and the use of retention time windows as an identification criterion, was identified as one of the most effective ways to improve selectivity. However, despite extensive optimisation of the chromatographic separation in Chapter 2, full base line resolution was not achieved for all target compounds. Additionally, the ability to separate non-target compounds will be dependent on the separation space available. There is a finite number of peaks that can be separated by a given method (peak capacity). One way to increase the separation space available, and reduce the probability of peaks overlapping, is the use of comprehensive two-dimensional (2D) chromatography. With comprehensive 2D chromatography the entire sample is separated in 2D, through two different stationary phases. If the two dimensions are orthogonal (i.e. there is no correlation in retention times) then the total 2D peak capacity is the product of the two individual peak capacities [148].

The use of comprehensive 2D gas chromatography, or GC x GC, in Forensic Science was reviewed in 2016 by Sampat et al. [149] and again in 2018 by Gruber et al. due to a sharp increase in the number of forensic studies using GC x GC [150]. Both reviews concluded that GC x GC offered the benefits of enhanced peak capacity and sensitivity in several forensic areas. The forensic areas identified by Sampat et al. included qualitive and quantitative trace analysis in complex matrices, such as toxicology samples, untargeted screening and identification in environmental forensics and chemical profiling of illicit drugs, ignitable liquids, CRBN agents and explosives [149]. In the more recent review, Gruber et al. identified a general trend towards the use of GC x GC for chemical profiling, fingerprinting and classification of evidence types with complex chemical compositions and an increase in the use of chemometrics [150]. Both reviews also recognised that further work was required before more widespread and routine forensic use of GC x GC. Sampat et al. recognised the need for easier data treatment and analysis, in addition to suitable databases to aid interpretation [149]. Similarly, Gruber et al. identified the need for standardised analytical methods and data interpretation [150], which would also be required for the development of databases.

Comprehensive 2D liquid chromatography, or LC x LC, has already shown promise for pharmaceutical analysis [151], but to the authors knowledge has not yet been used for the

analysis of explosives. Iguiniz and Heinisch acknowledged the attractiveness of the high peak capacity and concluded that since LC x LC has become more affordable the number of applications should increase. Therefore, future work could investigate the potential for LC x LC to improve peak capacity and increase confidence of identification in explosive screening. In addition to developing an analytical method, suitable methods for data analysis and interpretation would also be required.

References

[1] M López-López, C García-Ruiz, Infrared and Raman spectroscopy techniques applied to identification of explosives, TrAC Trends in Analytical Chemistry. 54 (2014) 36-44.

[2] S Singh, Sensors—An effective approach for the detection of explosives, Journal of Hazardous Materials. 144 (2007) 15-28.

[3] CPS, Explosives - Legal Guidance, https://www.cps.gov.uk/legal-guidance/explosives. 2018.

[4] Offences Against the Person Act 1861, (1861) Chapter 100 24-25 Vict.

[5] Criminal Damage Act 1971, (1971) Chapter 48.

[6] Poisons Act 1972, (1972) Chapter 66.

[7] The Poisons Act 1972 (Explosives Precursors) (Amendment) Regulations 2018, 415 (2018).

[8] Explosives Act 1875, (1875) Chapter 17 38-39 Vict.

[9] Criminal Law Act 1977, (1977) Chapter 45.

[10] Explosive Substances Act 1883, (1883) Chapter 3 46-47 Vict.

[11] Terrorism Act 2000, (2000) Chapter 11.

[12] The Fireworks Regulations 2004, 1836 (2004).

[13] The Explosives Regulations 2014, 1638 (2014).

[14] The Explosives Regulations 2014 (Amendment) Regulations 2016, 315 (2016).

[15] SM Kaye, HL Herman, Index for Encyclopedia Volumes 1 through 10, Encyclopedia of explosives and related items. Volume 10, United States Defense Technical Information Center, 1983, pp. A1-Z2.

[16] HSE, Penalties, 2018 Chapter 17 38-39 Vict.

[17] S Bell, Forensic chemistry, 2nd ed., Glenview, IL ; London : Pearson 2013.

[18] R Matyáš, A Lycka, R Jirásko, Z Jakový, J Maixner, L Mišková, et al. Analytical Characterization of Erythritol Tetranitrate, an Improvised Explosive, Journal of Forensic Sciences. 61 (2016) 759-764.

[19] N Rizzo, Erythritol 101: The Artificial Sweetener Of The Moment, http://womensrunning.competitor.com/2018/06/nutrition/erythritol-101-artificial-sweetener-moment_94122. (2018) 02/07/2018.

[20] B Vogelsanger. Chemical stability, compatibility and shelf life of explosives, CHIMIA. 58 (2004) 401-408.

[21] D Kalderis, AL Juhasz, R Boopathy, S Comfort. Soils contaminated with explosives: Environmental fate and evaluation of state-of-the-art remediation processes (IUPAC Technical Report), Pure Applied Chemistry. 83 (2011) 1407-1484.

[22] HA Yu, DA DeTata, SW Lewis, N Nic Daeid, The stability of TNT, RDX and PETN in simulated post-explosion soils: Implications of sample preparation for analysis, Talanta. 164 (2017) 716-726.

[23] D Trache, AF Tarchoun. Stabilizers for nitrate ester-based energetic materials and their mechanism of action: a state-of-the-art review, Journal of Materials Science. 53 (2018) 100-123.

[24] DM Badgujar, MB Talawar, SN Asthana, PP Mahulikar. Advances in science and technology of modern energetic materials: An overview, Journal of Hazardous Materials. 151 (2008) 289-305.

[25] TWGFEX Laboratory Explosion Group Standards And Protocols Committee, Recommended Guidelines for Forensic Identification of Intact Explosives, https://docs.wixstatic.com/ugd/4344b0_48a898605b19408d9e0b649afc5febbe.pdf. (2004).

[26] TWGFEX Laboratory Explosion Group Standards And Protocols Committee, Recommended Guidelines for Forensic Identification of Post-Blast Explosive Residues, https://docs.wixstatic.com/ugd/4344b0_8d6e256c0a2a415ea5b0f755db1551c0.pdf. (2007).

[27] ENFSI, Best Practice Manual for the Forensic Recovery, Identification and Analysis of Explosives Traces, (2015).

[28] E Sisco, J Dake, C Bridge. Screening for trace explosives by AccuTOF™-DART®: An in-depth validation study, Forensic Science International. 232 (2013) 160-168.

[29] OL Collin, C Niegel, KE DeRhodes, BR McCord, GP Jackson. Fast Gas Chromatography of Explosive Compounds Using a Pulsed-Discharge Electron Capture Detector, Journal of Forensic Sciences. 51 (2006) 815-818.

[30] E Holmgren, S Ek, A Colmsjö, Extraction of explosives from soil followed by gas chromatography–mass spectrometry analysis with negative chemical ionization, Journal of Chromatography A. 1222 (2012) 109-115.

[31] OL Collin, CM Zimmermann, GP Jackson. Fast gas chromatography negative chemical ionization tandem mass spectrometry of explosive compounds using dynamic collision-induced dissociation, International Journal of Mass Spectrometry. 279 (2009) 93-99.

[32] D DeTata, P Collins, A McKinley, A fast liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QToF-MS) method for the identification of organic explosives and propellants, Forensic Science International. 233 (2013) 63-74.

[33] X Xu, M Koeberg, C Kuijpers, E Kok. Development and validation of highly selective screening and confirmatory methods for the qualitative forensic analysis of organic explosive compounds with high performance liquid chromatography coupled with (photodiode array and) LTQ ion trap/Orbitrap mass spectrometric detections (HPLC-(PDA)-LTQOrbitrap), Science and Justice. 54 (2014) 3-21.

[34] X Zhao, J Yinon. Identification of nitrate ester explosives by liquid chromatography-electrospray ionization and atmospheric pressure chemical ionization mass spectrometry, Journal of Chromatography A. 977 (2002) 59-68.

[35] JA Mathis, BR McCord. The analysis of high explosives by liquid chromatography/electrospray ionization mass spectrometry: multiplexed detection of negative ion adducts, Rapid Communications in Mass Spectrometry. 19 (2005) 99-104.

[36] R Tachon, V Pichon, MBL Borgne, J Minet. Use of porous graphitic carbon for the analysis of nitrate ester, nitramine and nitroaromatic explosives and by-products by liquid chromatography-atmospheric pressure chemical ionisation-mass spectrometry, Journal of Chromatography A. 1154 (2007) 174-181.

[37] R Tachon, V Pichon, BL Borgne, J Minet. Comparison of solid-phase extraction sorbents for sample clean-up in the analysis of organic explosives, Journal of Chromatography A. 1185 (2008) 1-8.

[38] E Holmgren, H Carlsson, P Goede, C Crescenzi, Determination and characterization of organic explosives using porous graphitic carbon and liquid chromatography–atmospheric pressure chemical ionization mass spectrometry, Journal of Chromatography A. 1099 (2005) 127-135.

[39] JL Thomas, D Lincoln, BR McCord. Separation and Detection of Smokeless Powder Additives by Ultra Performance Liquid Chromatography with Tandem Mass Spectrometry (UPLC/MS/MS), Journal of Forensic Sciences. 58 (2013) 609-615.

[40] XM Xu, van de Craats, A. M., P de Bruyn. Highly sensitive screening method for nitroaromatic, nitramine and nitrate ester explosives by high performance liquid chromatography - Atmospheric pressure ionization - Mass spectrometry (HPLC-API-MS) in forensic applications, Journal of Forensic Sciences. 49 (2004) 1171-1180.

[41] L Song, JE Bartmess. Liquid chromatography/negative ion atmospheric pressure photoionization mass spectrometry: a highly sensitive method for the analysis of organic explosives, Rapid Communications in Mass Spectrometry. 23 (2009) 77-84.

[42] J Bečanová, Z Friedl, Z Šimek, Identification and determination of trinitrotoluenes and their degradation products using liquid chromatography–electrospray ionization mass spectrometry, International Journal of Mass Spectrometry. 291 (2010) 133-139.

[43] SA Oehrle. Analysis of explosives using ultra performance liquid chromatography (UPLC (R)) with UV and/or mass spectrometry detection, Journal of Energetic Materials. 26 (2008) 197-206.

[44] D Perret, S Marchese, A Gentili, R Curini, A Terracciano, E Bafile, et al. LC-MS-MS Determination of Stabilizers and Explosives Residues in Hand-Swabs, Chromatographia. 68 (2008) 517-524.

[45] U Ochsenbein, M Zeh, J Berset. Comparing solid phase extraction and direct injection for the analysis of ultra-trace levels of relevant explosives in lake water and tributaries using liquid chromatography-electrospray tandem mass spectrometry, Chemosphere. 72 (2008) 974-980.

[46] K Levsen, P Mussmann, E Bergerpreiss, A Preiss, D Volmer, G Wunsch. Analysis of nitroaromatics and nitramines in ammunition waste-water and in aqueous samples from former ammunition plants and other military sites, CLEAN Soil Air Water. 21 (1993) 153-166.

[47] DA DeTata, PA Collins, AJ McKinley. A Comparison of Common Swabbing Materials for the Recovery of Organic and Inorganic Explosive Residues, Journal of Forensic Sciences. 58 (2013) 757-763.

[48] X Liu, A Bordunov, M Tracy, C Pohl. A total solution to baseline separation of 14 explosives in US EPA method 8330, American Laboratory. 39 (2007) 28-30.

[49] T Borch, R Gerlach. Use of reversed-phase high-performance liquid chromatography-diode array detection for complete separation of 2,4,6-trinitrotoluene metabolites and EPA Method 8330 explosives: influence of temperature and an ion-pair reagent, Journal of Chromatography A. 1022 (2004) 83-94.

[50] S Lordel-Madeleine, V Eudes, V Pichon. Identification of the nitroaromatic explosives in post-blast samples by online solid phase extraction using molecularly imprinted silica sorbent coupled with reversed-phase chromatography, Analytical and Bioanalytical Chemistry. 405 (2013) 5237-5247.

[51] E Tyrrell, GW Dicinoski, EF Hilder, RA Shellie, MC Breadmore, CA Pohl, et al. Coupled reversed-phase and ion chromatographic system for the simultaneous identification of inorganic and organic explosives, Journal of Chromatography A. 1218 (2011) 3007-3012.

[52] B Paull, C Roux, M Dawson, P Doble. Rapid screening of selected organic explosives by high performance liquid chromatography using reversed-phase monolithic columns, Journal of Forensic Sciences. 49 (2004) 1181-1186.

[53] Gaurav, AK Malik, PK Rai, Development of a new SPME–HPLC–UV method for the analysis of nitro explosives on reverse phase amide column and application to analysis of aqueous samples, Journal of Hazardous Materials. 172 (2009) 1652-1658.

[54] WG Reifenrath, HO Kammen, WG Palmer, MM Major, GJ Leach. Percutaneous Absorption of Explosives and Related Compounds: An Empirical Model of Bioavailability of Organic Nitro Compounds from Soil, Toxicology and Applied Pharmacology. 182 (2002) 160-168.

[55] RL Marple, WR LaCourse. Application of Photoassisted Electrochemical Detection to Explosive-Containing Environmental Samples, Analytical Chemistry. 77 (2005) 6709-6714.

[56] D Gaurav, AK Malik, PK Rai. High-Performance Liquid Chromatographic Methods for the Analysis of Explosives, Critical Reviews in Analytical Chemistry. 37 (2007) 227-268.

[57] USEP Agency, Method 8330 - Nitroaromatics and nitramines by high performance liquid chromatography, (1994).

[58] USEP Agency, Method 8330A - Nitroaromatics and nitramines by high performance liquid chromatography, (2007).

[59] USEP Agency, Method 8330 B - Nitroaromatics, nitramines and nitrate esters by high performance liquid chromatography, (2006).

[60] P Kebarle. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, Journal of Mass Spectrometry. 35 (2000) 804-817.

[61] H Awad, MM Khamis, A El-Aneed. Mass Spectrometry, Review of the Basics: Ionization, Applied Spectroscopy Reviews. 50 (2015) 158-175.

[62] NB Cech, CG Enke. Practical implications of some recent studies in electrospray ionization fundamentals, Mass Spectrometry Reviews. 20 (2001) 362-387.

[63] CN McEwen, BS Larsen. Ionization Mechanisms Related to Negative Ion APPI, APCI, and DART, Journal of the American Society for Mass Spectrometry. 20 (2009) 1518-1521.

[64] XM Xu, van de Craats, A. M., EM Kok, P de Bruyn. Trace analysis of peroxide explosives by high performance liquid chromatography-atmospheric pressure chemical ionization-tandem mass spectrometry (HPLC-APCI-MS/MS) for forensic applications, Journal of Forensic Sciences. 49 (2004) 1230-1236.

[65] QZ Hu, RJ Noll, HY Li, A Makarov, M Hardman, RG Cooks. The Orbitrap: a new mass spectrometer, Journal of Mass Spectrometry. 40 (2005) 430-443.

[66] RA Zubarev, A Makarov. Orbitrap Mass Spectrometry, Analytical Chemistry. 85 (2013) 5288-5296.

[67] K Murray Kermit, K Boyd Robert, N Eberlin Marcos, GJ Langley, L Li, Y Naito, Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013), Pure Applied Chemistry. 85 (2013) 1515-1609.

[68] M Nic, J Jirat, B Kosata, IUPAC Gold Book, 2018 (2014).

[69] ThermoFisher Scientific, Can Your Q-TOF...Keep Pace with Your Needs? 2018.

[70] Waters, QuanTof: High-resolution, accurate mass, quantitative time-of-flight MS technology, 2018.

[71] M Krauss, H Singer, J Hollender. LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns, Analytical and Bioanalytical Chemistry. 397 (2010) 943-951.

[72] M Ibáñez, JV Sancho, L Bijlsma, ALN van Nuijs, A Covaci, F Hernández, Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances, TrAC Trends in Analytical Chemistry. 57 (2014) 107-117.

[73] R Put, Y Vander Heyden, Review on modelling aspects in reversed-phase liquid chromatographic quantitative structure–retention relationships, Analytica Chimica Acta. 602 (2007) 164-172.

[74] K Héberger, Quantitative structure–(chromatographic) retention relationships, Journal of Chromatography A. 1158 (2007) 273-305.

[75] R Kaliszan. QSRR: Quantitative structure-(chromatographic) retention relationships, Chemical reviews. 107 (2007) 3212-3246.

[76] RIJ Amos, PR Haddad, R Szucs, JW Dolan, CA Pohl, Molecular modeling and prediction accuracy in Quantitative Structure-Retention Relationship calculations for chromatography, TrAC Trends in Analytical Chemistry. 105 (2018) 352-359.

[77] K Goryński, B Bojko, A Nowaczyk, A Buciński, J Pawliszyn, R Kaliszan, Quantitative structure–retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: Endogenous metabolites and banned compounds, Analytica Chimica Acta. 797 (2013) 13-19.

[78] T Bruderer, E Varesio, G Hopfgartner, The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition, Journal of Chromatography B. 1071 (2017) 3-10.

[79] X Zhang, J Li, C Wang, D Song, C Hu, Identification of impurities in macrolides by liquid chromatography–mass spectrometric detection and prediction of retention times of impurities by constructing quantitative structure–retention relationship (QSRR), Journal of Pharmaceutical and Biomedical Analysis. 145 (2017) 262-272.

[80] R Aalizadeh, NS Thomaidis, AA Bletsou, P Gago-Ferrero. Quantitative Structure-Retention Relationship Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples, Journal of Chemical Information and Modeling. 56 (2016) 1384-1398.

[81] M Goodarzi, R Jensen, Y Vander Heyden, QSRR modeling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions, Journal of Chromatography B. 910 (2012) 84-94.

[82] TH Miller, A Musenga, DA Cowan, LP Barron. Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks, Analytical Chemistry. 85 (2013) 10330-10337.

[83] K Munro, TH Miller, CPB Martins, AM Edge, DA Cowan, LP Barron, Artificial neural network modelling of pharmaceutical residue retention times in wastewater extracts using gradient liquid chromatography-high resolution mass spectrometry data, Journal of Chromatography A. 1396 (2015) 34-44.

[84] LP Barron, GL McEneff. Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across 10 multi-residue reversed-phase analytical methods, Talanta. 147 (2016) 261-270.

[85] DJ Livingstone, DT Manallack, IV Tetko. Data modelling with neural networks: Advantages and limitations, Journal of computer-aided molecular design. 11 (1997) 135-142.

[86] Trajan, Trajan 6.0 Professional Neural Network Simulator, 2001.

[87] S Yousefinejad, B Hemmateenejad. Chemometrics tools in QSAR/QSPR studies: A historical perspective, Chemometrics and Intelligent Laboratory Systems. 149 (2015) 177-204.

[88] H Noorizadeh, A Farmany, H Narimani, M Noorizadeh. QSRR using evolved artificial neural network for 52 common pharmaceuticals and drugs of abuse in hair from UPLC-TOF-MS, Drug Testing and Analysis. 5 (2013) 320-324.

[89] CB Mollerup, M Mardal, PW Dalsgaard, K Linnet, LP Barron, Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry, Journal of Chromatography A. 1542 (2018) 82-88.

[90] R Bade, L Bijlsma, TH Miller, LP Barron, JV Sancho, F Hernández. Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis, Science of the Total Environment. 538 (2015) 934-941.

[91] P Žuvela, K Macur, J Jay Liu, T Bączek, Exploiting non-linear relationships between retention time and molecular structure of peptides originating from proteomes and comparing three multivariate approaches, Journal of Pharmaceutical and Biomedical Analysis. 127 (2016) 94-100.

[92] B Rochat. Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data, Journal of the American Society for Mass Spectrometry. 28 (2017) 709-723.

[93] European Commission, Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, 2002/657/EC (2002).

[94] SJ Lehotay, K Mastovska, A Amirav, AB Fialkov, PA Martos, Ad Kok, et al., Identification and confirmation of chemical residues in food by chromatography-mass spectrometry and other techniques, TrAC Trends in Analytical Chemistry. 27 (2008) 1070-1090.

[95] A Crowson, MS Beardah. Development of an LC/MS method for the trace analysis of hexamethylenetriperoxidediamine (HMTD), Analyst. 126 (2001) 1689-1693.

[96] L Widmer, S Watson, K Schlatter, A Crowson. Development of an LC/MS method for the trace analysis of triacetone triperoxide (TATP), Analyst. 127 (2002) 1627-1632.

[97] MJ Lang, SE Burns. Improvement of EPA method 8330: complete separation using a two-phase approach, Journal of Chromatography A. 849 (1999) 381-388.

[98] VG Sears, SM Bleay, HL Bandey, VJ Bowman. A methodology for finger mark research, Science and Justice. 52 (2012) 145-160.

[99] Agilent Technologies, Understanding Your ChemStation, Agilent ChemStation. (2009) 240-241.

[100] L Salvia, The investigation and evaluation of stationary phases for high performance reverse phase liquid chromatography of nitroaromatic, nitramine, nitroaliphatic and nitrate ester based organic explosives, Unpublished MSc Thesis, King's College London (2013).

[101] Ion Max and Ion Max-S API Source Hardware Manual, Revision C (2007).

[102] Exactive™ Software Manual, Revision C - 1249910 (2011).

[103] F Hernández, M Ibáñez, R Bade, L Bijlsma, JV Sancho, Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry, TrAC Trends in Analytical Chemistry. 63 (2014) 140-157.

[104] HGJ Mol, P Zomer, M De Koning. Qualitative aspects and validation of a screening method for pesticides in vegetables and fruits based on liquid chromatography coupled to full scan high resolution (Orbitrap) mass spectrometry, Analytical and Bioanalytical Chemistry. 403 (2012) 2891-2908.

[105] Y Fu, C Zhao, X Lu, G Xu, Nontargeted screening of chemical contaminants and illegal additives in food based on liquid chromatography–high resolution mass spectrometry, TrAC Trends in Analytical Chemistry. 96 (2017) 89-98.

[106] D Pasin, A Cawley, S Bidny, S Fu. Current applications of high-resolution mass spectrometry for the analysis of new psychoactive substances: a critical review, Analytical and Bioanalytical Chemistry. 409 (2017) 5821-5836.

[107] European Commission, Guidance document on analytical quality control and method validation procedures for pesticide residues and analysis in food and feed. Official Journal of the European Communities. SANTE/11813/2017 (2017).

[108] Validation of analytical procedures: Text and methodology, ICH Harmonised Tripartite Guideline. Q2(R1) (2005).

[109] GL McEneff, B Murphy, T Webb, D Wood, R Irlam, J Mills, et al. Sorbent Film-Coated Passive Samplers for Explosives Vapour Detection Part A: Materials Optimisation and Integration with Analytical Technologies, Scientific Reports. 8 (2018).

[110] JT Watson, OD Sparkman, Introduction to Mass Spectrometry : Instrumentation, Applications, and Strategies for Data Interpretation, Wiley, Chichester, 2008, pp. 271-272.

[111] J Florián, L Gao, V Zhukhovskyy, DK MacMillan, MP Chiarelli, Nitramine Anion Fragmentation: A Mass Spectrometric and Ab Initio Study, Journal of the American Society for Mass Spectrometry. 18 (2007) 835-841.

[112] L Havlikova, R Matyas, L Ihnat, L Novakova, D Satinsky. Degradation study of nitroaromatic explosives 2-diazo-4,6-dinitrophenol and picramic acid using HPLC and UHPLC-ESI-MS/MS, Analytical Methods. 6 (2014) 4761-4768.

[113] B Nguyen Van, EV Nikolaeva, AG Shamov, GM Khrapkovskii, RV Tsyshevsky, Exploration of decomposition pathways of 2,4,6-trinitrotoluene (TNT) radical ions by means of density functional theory, International Journal of Mass Spectrometry. 392 (2015) 7-15.

[114] A Ostrinskaya, JA Kelley, RR Kunz. Characterization of nitrated sugar alcohols by atmospheric-pressure chemical-ionization mass spectrometry, Rapid Communications in Mass Spectrometry. 31 (2017) 333-343.

[115] K Colizza, KE Mahoney, AV Yevdokimov, JL Smith, JC Oxley. Acetonitrile Ion Suppression in Atmospheric Pressure Ionization Mass Spectrometry, Journal of the American Society for Mass Spectrometry. 27 (2016) 1796-1804.

[116] K Colizza, A Yevdokimov, L McLennan, JL Smith, JC Oxley. Reactions of Organic Peroxides with Alcohols in Atmospheric Pressure Chemical Ionization-the Pitfalls of Quantifying Triacetone Triperoxide (TATP), Journal of the American Society for Mass Spectrometry. 29 (2018) 393-404.

[117] Royal Society of Chemistry, ChemSpider, http://www.chemspider.com/AboutUs.aspx. 22/03/18.

[118] H Horai, M Arita, S Kanaya, Y Nihei, T Ikeda, K Suwa, et al. MassBank: A public repository for sharing mass spectral data for life sciences, Journal of Mass Spectrometry. 45 (2010) 703-714.

[119] JL Little, AJ Williams, A Pshenichnov, V Tkachenko. Identification of "known unknowns" utilizing accurate mass data and chemspider, Journal of the American Society for Mass Spectrometry. 23 (2012) 179-185.

[120] H Brust, S Willemse, T Zeng, A van Asten, M Koeberg, A van der Heijden, et al., Impurity profiling of trinitrotoluene using vacuum-outlet gas chromatography–mass spectrometry, Journal of Chromatography A. 1374 (2014) 224-230.

[121] RG Ewing, MJ Waltman, DA Atkinson, JW Grate, PJ Hotchkiss, The vapor pressures of explosives, TrAC Trends in Analytical Chemistry. 42 (2013) 35-48.

[122] M Howard, The Marking of Plastic Explosives for Detection Regulations 1996, 890 (1996).

[123] S Nacson, A Grigoriev, Transport and migration of explosive traces through airborne dust particles, International Symposium on the Analysis and Detection of Explosives. (2017).

[124] CA Arenal, BE Sample, Chapter 25 - Wildlife Toxicity Assessment for Diphenylamine, in: Williams MA, , Reddy G, , Quinn MJ, , et al. (Eds.), Wildlife Toxicity Assessments for Chemicals of Military Concern, Elsevier, 2015, pp. 439-464.

[125] ISO/CASCO, General requirements for the competence of testing and calibration laboratories, ISO/IEC 17025:2017. 3 (2017).

[126] T De Baere, W Dmitruk, D Meuwly, G O'Donnell, Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science, ENFSI. Version 2.0 (2014).

[127] M Thompson, SLR Ellison, R Wood. Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report), Pure and Applied Chemistry,. 74 (2002) 835-885.

[128] B Magnusson, U Örnemark, The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, Eurachem Guide. 2nd edition (2014).

[129] B Dejaegher, YV Heyden, Ruggedness and robustness testing, Journal of Chromatography A. 1158 (2007) 138-157.

[130] H Rapp-Wright, G McEneff, B Murphy, S Gamble, R Morgan, M Beardah, et al., Suspect screening and quantification of trace organic explosives in wastewater using solid phase extraction and liquid chromatography-high resolution accurate mass spectrometry, Journal of Hazardous Materials. 329 (2017) 11-21.

[131] S Kim, PA Thiessen, EE Bolton, J Chen, G Fu, A Gindulyte, et al. PubChem substance and compound databases, Nucleic acids research. 44 (2016) D1202-D1213.

[132] I Advanced Chemistry Development, ACD/ChemSketch (freeware), www.acdlabs.com. version 2015.2.5 (2015).

[133] IV Tetko, J Gasteiger, R Todeschini, A Mauri, D Livingstone, P Ertl, et al. Virtual computational chemistry laboratory - Design and description, Journal of computer-aided molecular design. 19 (2005) 453-463.

[134] M Kuhn, K Johnson, Applied Predictive Modeling, 1st ed., Springer-Verlag New York 2013.

[135] AA Petrauskas, EA Kolovanov. ACD/Log P method description, Perspectives in Drug Discovery and Design. 19 (2000) 99-116.

[136] AK Ghose, GM Crippen. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure-Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions, Journal of Chemical Information and Modeling. 27 (1987) 21-35.

[137] I Moriguchi, S Hirono, Q Liu, I Nakagome, Y Matsushita. Simple Method of Calculating Octanol/Water Partition Coefficient, Chemical and Pharmaceutical Bulletin. 40 (1992) 127-130.

[138] R Mannhold, GI Poda, C Ostermann, IV Tetko. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds, Journal of pharmaceutical sciences. 98 (2009) 861-893.

[139] LH Hall, B Mohney, LB Kier. The Electrotopological State: An Atom Index for QSAR, Quantitative Structure-Activity Relationships. 10 (1991) 43-51.

[140] LK Hansen, P Salamon. Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence. 12 (1990) 993-1001.

[141] MH Kutner, CJ Nachtsheim, J Neter, 2.5 Prediction of New Observation, Applied Linear Regression Models, 4th ed., McGraw-Hill Companies, 2004, pp. 55-61.

[142] DM Lane, Standard Error of the Estimate, Online Statistics Education: A Multimedia Course of Study, Rice University, http://onlinestatbook.com/Online_Statistics_Education.pdf, Accessed 07/06/18, pp. 473-475.

[143] Statistics How To, T-Distribution Table (One Tail and Two-Tails), http://www.statisticshowto.com/tables/t-distribution-table/. 2018 (2017).

[144] T Value Table, http://www.ttable.org/. 2018.

[145] H Oberacher, K Arnhard, Current status of non-targeted liquid chromatography-tandem mass spectrometry in forensic toxicology, TrAC Trends in Analytical Chemistry. 84 (2016) 94-105.

[146] K Arnhard, A Gottschall, F Pitterl, H Oberacher. Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry, Analytical and Bioanalytical Chemistry. 407 (2015) 405-414.

[147] S Broecker, S Herre, B Wuest, J Zweigenbaum, F Pragst. Development and practical application of a library of CID accurate mass spectra of more than 2,500 toxic compounds for systematic toxicological analysis by LC-QTOF-MS with data-dependent acquisition, Analytical and Bioanalytical Chemistry. 400 (2011) 101-117.

[148] P Dugo, F Cacciola, T Kumm, G Dugo, L Mondello, Comprehensive multidimensional liquid chromatography: Theory and applications, Journal of Chromatography A. 1184 (2008) 353-368.

[149] A Sampat, M Lopatka, M Sjerps, G Vivo-Truyols, P Schoenmakers, A van Asten, Forensic potential of comprehensive two-dimensional gas chromatography, TrAC Trends in Analytical Chemistry. 80 (2016) 345-363.

[150] B Gruber, BA Weggler, R Jaramillo, KA Murrell, PK Piotrowski, FL Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, TrAC Trends in Analytical Chemistry. 105 (2018) 292-301.

[151] M Iguiniz, S Heinisch, Two-dimensional liquid chromatography in pharmaceutical analysis. Instrumental aspects, trends and applications, Journal of Pharmaceutical and Biomedical Analysis. 145 (2017) 482-503.

Appendix

List of Tables

TABLE A. 1: LIST OF ANALYTES AND SMILES USED TO GENERATE MOLECULAR DESCRIPTORS..... 232

TABLE A. 2: MOLECULAR DESCRIPTORS AND PREDICTED RETENTION TIMES FOR ANN AND MLR MODELS USING THE GORYŃSKI ET AL. DESCRIPTORS. 239

TABLE A. 3: MOLECULAR DESCRIPTORS AND PREDICTED RETENTION TIME FOR BEST ANN MODEL USING THE BARRON & MCENEFF DESCRIPTORS. 245

TABLE A. 4: MOLECULAR DESCRIPTORS AND PREDICTED RETENTION TIMES FOR ANN AND MLR MODELS USING AALIZADEH ET AL. DESCRIPTORS 251

TABLE A. 5: MOLECULAR DESCRIPTORS AND PREDICTED RETENTION TIMES FOR BEST NETWORK AND ENSEMBLE FOLLOWING GENETIC ALGORITHM FEATURE SELECTION. 257

TABLE A. 6: MOLECULAR DESCRIPTORS AND PREDICTED RETENTION TIMES FOR BEST ANN AND ENSEMBLE FOLLOWING FORWARD SELECTION - PART A 268

TABLE A. 7: MOLECULAR DESCRIPTORS AND PREDICTED RETENTION TIMES FOR BEST ANN AND ENSEMBLE FOLLOWING FORWARD SELECTION - PART B 280

Table A. 1: List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
1,2-DNB	c1ccc(c(c1)[N+](=O)[O-])[N+](=O)[O-]
1,2-DNG	C(C(CO[N+](=O)[O-])O[N+](=O)[O-])O
(1E,2E,3E,4E,5E,6E)-N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	C1(=NO)C(=NO)C(=NO)C(=NO)C(=NO)C1=NO
(1Z,2Z,3E,4Z,5Z,6E)-N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	C1(=NO)C(=NO)C(=NO)C(=NO)C(=NO)C1=NO
1-(2-Methyl-3-nitro-2-propen-1-yl)-3,5-dinitro-1H-1,2,4-triazole	CC(=C[N+](=O)[O-])Cn1c(nc(n1)[N+](=O)[O-])[N+](=O)[O-]
1,2,2-Trinitro-1,3,5-triazinane	C1NCN(C(N1)([N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
1,3,5-triazinane-2,4,6-trione	c1(=O)[nH]c(=O)[nH]c(=O)[nH]1.c1(=O)[nH]c(=O)[nH]c(=O)[nH]1
1,3,5-Trinitro-1,2,3-triazinane	C1C(CN(NN1N(=O)=O)N(=O)=O)N(=O)=O
1,3-DNB	c1cc(cc(c1)[N+](=O)[O-])[N+](=O)[O-]
1,3-DNG	C(C(CO[N+](=O)[O-])O)O[N+](=O)[O-]
1-[(2Z)-2-Methyl-3-nitro-2-propen-1-yl]-3,5-dinitro-1H-1,2,4-triazole	C/C(=C/[N+](=O)[O-])/Cn1c(nc(n1)[N+](=O)[O-])[N+](=O)[O-]
1-MNG	C(C(CO[N+](=O)[O-])O)O
2-(N'-Nitro-N-nitroso-carbamimidamido)ethyl nitrate	O=[N+]([O-])OCCN(N=O)C(=N)N[N+]([O-])=O
2,4,6-Trinitro-1,3,5-triazinane	C1(NC(NC(N1)[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
2,4-DA-6-NT	Cc1c(cc(cc1[N+](=O)[O-])N)N
2,4-DNDPA	c1ccc(cc1)Nc2ccc(cc2[N+](=O)[O-])[N+](=O)[O-]
2,4-DNT	Cc1ccc(cc1[N+](=O)[O-])[N+](=O)[O-]
2,6-DA-4-NT	Cc1c(cc(cc1N)[N+](=O)[O-])N
2,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-3,5-pyridinediamine	O=[N+]([O-])c2nc(c(cc2Nc1c(cc(cc1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O)Nc3c(cc(cc3[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O
2,6-DNT	Cc1c(cccc1[N+](=O)[O-])[N+](=O)[O-]
2-A-4,6-DNT	Cc1c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])N
2-NDPA	c1ccc(cc1)Nc2ccccc2[N+](=O)[O-]
2-NT	Cc1ccccc1[N+](=O)[O-]
3,4-DNT	Cc1ccc(c(c1)[N+](=O)[O-])[N+](=O)[O-]
3,5-DNA	c1c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])N
3,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-2,5-pyridinediamine	[O-][N+](=O)c3cc(cc([N+]([O-])=O)c3Nc2cc(c(Nc1c(cc(cc1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O)nc2[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O
3-NT	Cc1cccc(c1)[N+](=O)[O-]
4,4'-DNDPA	c1cc(ccc1Nc2ccc(cc2)[N+](=O)[O-])[N+](=O)[O-]
4,5,6-Trinitro-1,2,3-benzenetriamine	c1(c(c(c(c(c1N)[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])N)N
4-A-2,6-DNT	Cc1c(cc(cc1[N+](=O)[O-])N)[N+](=O)[O-]
4-NDPA	c1ccc(cc1)Nc2ccc(cc2)[N+](=O)[O-]
4-NT	Cc1ccc(cc1)[N+](=O)[O-]
6-MAM	CC(=O)O[C@H]1C=C[C@H]2[C@H]3Cc4ccc(c5c4[C@]2([C@H]1O5)CCN3C)O
Acebutolol	CCCC(=O)Nc1ccc(c(c1)C(=O)C)OCC(CNC(C)C)O
Acetazolamide	CC(=O)Nc1nnc(s1)S(=O)(=O)N
AICAR	c1nc(c(n1[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O)N)C(=O)N
Alprenolol	CC(C)NCC(COc1ccccc1CC=C)O
Altizide	C=CCSC[C@@H]1Nc2cc(c(cc2S(=O)(=O)N1)S(=O)(=O)N)Cl
Amiloride	c1(c(nc(c(n1)Cl)N)N)/C(=N/C(=N)N)/O
Aminoglutethimide	CCC1(CCC(=O)NC1=O)c2ccc(cc2)N
Anastrazole	CC(C)(C#N)c1cc(cc(c1)C(C)(C)C#N)Cn2cncn2

Table A. 1 (Continued): List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
Andarine	CC(=O)NC1=CC=C(C=C1)OC[C@@](C)(C(=O)NC2=CC(=C(C=C2)[N+](=O)[O-])C(F)(F)F)O
Atenolol	CC(C)NCC(COc1ccc(cc1)CC(=O)N)O
Bambuterol	CC(C)(C)NCC(c1cc(cc(c1)OC(=O)N(C)C)OC(=O)N(C)C)O
Beclomethasone	C[C@H]1C[C@H]2[C@@H]3CCC4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CO)O)C)O)Cl)C
Bendroflumethiazide	c1ccc(cc1)CC2Nc3cc(c(cc3S(=O)(=O)N2)S(=O)(=O)N)C(F)(F)F
Benzoylecgonine	CN1[C@H]2CC[C@@H]1[C@H]([C@H](C2)OC(=O)c3ccccc3)C(=O)O
Benzthiazide	c1ccc(cc1)CSCC2=Nc3cc(c(cc3S(=O)(=O)N2)S(=O)(=O)N)Cl
Beta\dexameth.	C[C@H]1C[C@H]2[C@@H]3CCC4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CO)O)C)O)F)C
Betaxolol	CC(C)NCC(COc1ccc(cc1)CCOCC2CC2)O
Bisoprolol	CC(C)NCC(COc1ccc(cc1)COCCOC(C)C)O
Budesonide	CCCC1O[C@@H]2C[C@H]3[C@@H]4CCC5=CC(=O)C=C[C@@]5([C@H]4[C@H](C[C@@]3([C@@]2(O1)C(=O)CO)C)O)C
Bumetanide	CCCCNc1cc(cc(c1Oc2ccccc2)S(=O)(=O)N)C(=O)O
Bumetanide	CCCCNc1cc(cc(c1Oc2ccccc2)S(=O)(=O)N)C(=O)O
Bunolol	CC(C)(C)NC[C@@H](COc1cccc2c1CCCC2=O)O
Bupropion	CC(C(=O)c1cccc(c1)Cl)NC(C)(C)C
Butizide	CC(C)CC1Nc2cc(c(cc2S(=O)(=O)N1)S(=O)(=O)N)Cl
Canrenone	C[C@]12CCC(=O)C=C1C=C[C@@H]3[C@@H]2CC[C@]4([C@H]3CC[C@@]45CCC(=O)O5)C
Carphedon	c1ccc(cc1)C2CC(=O)N(C2)CC(=O)N
Carteolol	CC(C)(C)NCC(COc1cccc2c1CCC(=O)N2)O
Carvedilol	COc1ccccc1OCCNCC(COc2cccc3c2c4ccccc4[nH]3)O
Celiprolol	CCN(CC)C(=O)Nc1ccc(c(c1)C(=O)C)OCC(CNC(C)(C)C)O
Chlorexolone	c1c2c(cc(c1Cl)S(=O)(=O)N)C(=O)N(C2)C3CCCCC3
Chlorothiazide	c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)N=CN2
Chlorthalidone	c1ccc2c(c1)C(=O)NC2(c3ccc(c(c3)S(=O)(=O)N)Cl)O
Clobenzorex	CC(Cc1ccccc1)NCc2ccccc2Cl
Clobetasol	C[C@H]1C[C@H]2[C@@H]3CCC4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CCl)O)C)O)F)C
Clopamide	CC1CCCC(N1NC(=O)c2ccc(c(c2)S(=O)(=O)N)Cl)C
Cyclopenthiazide	c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)NC(N2)CC3CCCC3
DADP	CC1(OOC(OO1)(C)C)C
Danazol	C[C@]12CC[C@H]3[C@H]([C@@H]1CC[C@]2(C#C)O)CCC4=Cc5c(cno5)C[C@]34C
DEDPU	CCN(c1ccccc1)C(=O)N(CC)c2ccccc2
Deflazacort	CC1=N[C@@]2([C@H](O1)C[C@@H]3[C@@]2(C[C@@H]([C@H]4[C@H]3CCC5=CC(=O)C=C[C@]45C)O)C)C(=O)COC(=O)C
Desacetyl deflazacort	CC1=N[C@@]2([C@H](O1)C[C@@H]3[C@@]2(C[C@@H]([C@H]4[C@H]3CCC5=CC(=O)C=C[C@]45C)O)C)C(=O)CO
Desonide	C[C@]12C[C@@H]([C@H]3[C@H]([C@@H]1C[C@@H]4[C@]2(OC(O4)(C)C)C(=O)CO)CCC5=CC(=O)C=C[C@]35C)O
Diacetolol	CC(C)NCC(COc1ccc(cc1C(=O)C)NC(=O)C)O
Dichlorphenamide	c1c(cc(c(c1S(=O)(=O)N)Cl)Cl)S(=O)(=O)N
Dimethyl phthalate	COC(=O)c1ccccc1C(=O)OC
DMDPU	CN(c1ccccc1)C(=O)N(C)c2ccccc2
DMNB	CC(C)(C(C)(C)[N+](=O)[O-])[N+](=O)[O-]
DPA	c1ccc(cc1)Nc2ccccc2
Efaproxiral	Cc1cc(cc(c1)NC(=O)Cc2ccc(cc2)OC(C)(C)C(=O)O)C

Table A. 1 (Continued): List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
EGDN	C(CO[N+](=O)[O-])O[N+](=O)[O-]
Epitizide	c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)NC(N2)CSCC(F)(F)F
Esmolol	CC(C)NCC(COc1ccc(cc1)CCC(=O)OC)O
Etacr. acid	CCC(=C)C(=O)c1ccc(c(c1Cl)Cl)OCC(=O)O
Etamivan	CCN(CC)C(=O)c1ccc(c(c1)OC)O
ETN	C([C@H]([C@H](CO[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-]
Famprofazone	CC(C)C1=C(N(N(C1=O)C2=CC=CC=C2)C)CN(C)C(C)CC3=CC=CC=C3
Fenbutrazate	CCC(c1ccccc1)C(=O)OCCN2CCOC(C2C)c3ccccc3
Fencamfamin	CCNC1C2CCC(C2)C1c3ccccc3
Fenetylline	CC(CC1=CC=CC=C1)NCCN2C=NC3=C2C(=O)N(C(=O)N3C)C
Fenfluramine	CCNC(C)Cc1cccc(c1)C(F)(F)F
Fludrocortisone	C[C@]12CCC(=O)C=C1CC[C@@H]3[C@@]2([C@H](C[C@]4([C@H]3CC[C@@]4(C(=O)CO)O)C)O)F
Flumethasone	C[C@@H]1C[C@H]2[C@@H]3C[C@@H](C4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CO)O)C)O)F)C)F
Fluticasone prop.	CCC(=O)O[C@@]1([C@@H](C[C@@H]2[C@@]1(C[C@@H]([C@]3([C@H]2C[C@@H](C4=CC(=O)C=C[C@@]43C)F)F)O)C)C)C(=O)SCF
Formoterol	C[C@H](Cc1ccc(cc1)OC)NC[C@@H](c2ccc(c(c2)NC=O)O)O
FPCAM	O=C(O[C@@]3(C(=O)O)[C@]2(C[C@H](O)[C@]4(F)[C@@]/1(\C(=C/C(=O)\C=C\1)[C@@H](F)C[C@H]4[C@@H]2C[C@H]3C)C)C)CC
Fulvestrant	C[C@]12CC[C@@H]3c4ccc(cc4C[C@H]([C@H]3[C@@H]1CC[C@@H]2O)CCCCCCCCCS(=O)CCCC(C(F)(F)F)(F)F)O
Furosemide	c1cc(oc1)CNc2cc(c(cc2C(=O)O)S(=O)(=O)N)Cl
Gestrinone	CCC12C=CC3=C4CCC(=O)C=C4CCC3C1CCC2(C#C)O
HMDD	C1CN2COOCN1COOC2
HMTA	C1N2CN3CN1CN(C2)C3
HMTD	C1N2COOCN(COO1)COOC2
HMX	C1N(CN(CN(CN1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
HND	c1c(cc(c(c1[N+](=O)[O-])Nc2c(cc(cc2[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
Hydrochloroth.	c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)NCN2
Hydroflumeth.	c1c(c(cc2c1NCNS2(=O)=O)S(=O)(=O)N)C(F)(F)F
Isometheptene	CC(CCC=C(C)C)NC
Labetalol	CC(CCc1ccccc1)NCC(c2ccc(c(c2)C(=O)N)O)O
M-1,2,3,4,5-PN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](CO)O[N+]([O-])=O
M-1,2,3,4,6-PN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
M-1,2,3,4-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO
M-1,2,3,5,6-PN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO[N+]([O-])=O)O[N+]([O-])=O
M-1,2,3,5-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO)O[N+]([O-])=O
M-1,2,3,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](O)CO[N+]([O-])=O
M-1,2,3-TriN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](O)CO
M-1,2,4,5-TN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO[N+]([O-])=O)O[N+]([O-])=O
M-1,2,4,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
M-1,2,4-TriN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O[N+]([O-])=O)[C@H](O)CO

Table A. 1 (Continued): List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
M-1,2,5,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@@H](CO[N+]([O-])=O)O[N+]([O-])=O
M-1,2,6-TriN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@H](O)CO[N+]([O-])=O
M-1,3,4,5-TN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
M-1,3,4,6-TN	O=[N+]([O-])O[C@@H]([C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
M-1,3,4-TriN	O=[N+]([O-])O[C@@H]([C@H](O[N+]([O-])=O)[C@H](O)CO)[C@H](O)CO[N+]([O-])=O
M-1,3,6-TriN	O[C@@H]([C@H](O[N+](=O)[O-])[C@H](O)CO[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
M-2,3,4,5-TN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](CO)O[N+]([O-])=O
M-2,3,4-TriN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO
Mefenorex	CC(Cc1ccccc1)NCCCCl
Mefruside	CC1(CCCO1)CN(C)S(=O)(=O)c2ccc(c(c2)S(=O)(=O)N)Cl
MEKP CP dimer	CCC1(OOC(OO1)(C)CC)C
MEKP CP tetramer	CCC1(C)OOC(C)(OOC(C)(OOC(C)(OO1)CC)CC)CC
MEKP CP trimer	CCC1(OOC(OOC(OO1)(C)CC)(C)CC)C
MEKP DHP dimer	CCC(C)(OO)OOC(C)(CC)OO
MEKP DHP monomer	CCC(C)(OO)OO
MEKP DHP pentamer	CCC(C)(OO)OOC(C)(CC)OOC(C)(CC)OOC(C)(CC)OOC(C)(CC)OO
MEKP DHP tetetramer	CCC(C)(OO)OOC(C)(CC)OOC(C)(CC)OOC(C)(CC)OO
MEKP DHP trimer	CCC(C)(OO)OOC(C)(CC)OOC(C)(CC)OO
Methylprednisolone	C[C@H]1C[C@H]2[C@@H]3CC[C@@]([C@]3(C[C@@H]([C@@H]2[C@@]4(C1=CC(=O)C=C4)C)O)C)(C(=O)CO)O
Metoprolol	CC(C)NCC(COc1ccc(cc1)CCOC)O
MHN	C([C@H]([C@H]([C@@H]([C@@H](CO[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-]
N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	C1(=NO)C(=NO)C(=NO)C(=NO)C(=NO)C1=NO
N,N',N''-Trihydroxy-2,4,6-trinitroso-1,3,5-benzenetriamine	c1(c(c(c(c(c1N=O)NO)N=O)NO)N=O)NO
N,N',N''-Trinitro-1,2,3-cyclopropanetriamine	O=[N+]([O-])NC1C(N[N+]([O-])=O)C1N[N+](=O)[O-]
N-{1-[Amino(oxo)acetyl]-3,5-dioxo-1,2,4-triazolidin-4-yl}ethanediamide	O=C1N(NC(=O)C(N)=O)C(=O)NN1C(=O)C(N)=O
Nadoxolol	c1ccc2c(c1)cccc2OCC(C/C(=N/O)/N)O
NB	c1ccc(cc1)[N+](=O)[O-]
NG	C(C(CO[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-]
Nikethamide	CCN(CC)C(=O)c1cccnc1
NM	C[N+](=O)[O-]
N-nitroso-DPA	c1ccc(cc1)N(c2ccccc2)N=O
Norfenfluramine	CC(Cc1cccc(c1)C(F)(F)F)N
NQ	C(=N)(N)N[N+](=O)[O-]
Ostarine	CC(=O)NC1=CC=C(C=C1)OC[C@@](C)(C(=O)NC2=CC(=C(C=C2)[N+](=O)[O-])C(F)(F)F)O
Oxprenolol	CC(C)NCC(COc1ccccc1OCC=C)O
Oxycodone	CN1CC[C@]23c4c5ccc(c4O[C@H]2C(=O)CC[C@]3([C@H]1C5)O)OC
Oxymorphone	C[N@]1CC[C@]23c4c5ccc(c4O[C@H]2C(=O)CC[C@]3([C@H]1C5)O)O
PA	c1c(cc(c(c1[N+](=O)[O-])O)[N+](=O)[O-])[N+](=O)[O-]
Pemoline	c1ccc(cc1)C2C(=O)N=C(O2)N
Pentazocine	C[C@H]1[C@H]2Cc3ccc(cc3[C@@]1(CCN2CC=C(C)C)C)O

Table A. 1 (Continued): List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
Pentetrazol	C1CCc2nnnn2CC1
Pethidine	CCOC(=O)C1(CCN(CC1)C)c2ccccc2
PETN	C(C(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-])O[N+](=O)[O-]
PGDN	CC(CO[N+](=O)[O-])O[N+](=O)[O-]
Phendimetrazine	CC1C(OCCN1C)c2ccccc2
Phenmetrazine	CC1C(OCCN1)c2ccccc2
Pindolol	CC(C)NCC(COc1cccc2c1cc[nH]2)O
Piretanide	c1ccc(cc1)Oc2c(cc(cc2S(=O)(=O)N)C(=O)O)N3CCCC3
Polythiazide	CN1C(Nc2cc(c(cc2S1(=O)=O)S(=O)(=O)N)Cl)CSCC(F)(F)F
Prenylamine	CC(Cc1ccccc1)NCCC(c2ccccc2)c3ccccc3
Probenecid	CCCN(CCC)S(=O)(=O)c1ccc(cc1)C(=O)O
Prolintane	CCCC(Cc1ccccc1)N2CCCC2
Propranolol	CC(C)NCC(COc1cccc2c1cccc2)O
Propylhexedrine	CC(CC1CCCCC1)NC
PYX	C1=C(C=C(C(=C1[N+](=O)[O-])NC2=C(C=C(C(=N2)NC3=C(C=C(C=C3[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
RDX	C1N(CN(CN1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
R-Salt	C1N(CN(CN1N=O)N=O)N=O
S-1,2,3,4,5-PN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](CO)O[N+]([O-])=O
S-1,2,3,4,6-PN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](O)CO[N+]([O-])=O
S-1,2,3,4-TN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO
S-1,2,3,5,6-PN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](CO[N+]([O-])=O)O[N+]([O-])=O
S-1,2,3,5-TN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO)O[N+]([O-])=O
S-1,2,3,6-TN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](O)CO[N+]([O-])=O
S-1,2,3-TrN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](O)CO
S-1,2,4,5,6-PN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO[N+]([O-])=O)O[N+]([O-])=O
S-1,2,4,5-TN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](CO[N+]([O-])=O)O[N+]([O-])=O
S-1,2,4,6-TN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
S-1,2,4-TrN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O[N+]([O-])=O)[C@H](O)CO
S-1,2,5,6-TN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@@H](CO[N+]([O-])=O)O[N+]([O-])=O
S-1,2,5-TrN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@@H](CO)O[N+]([O-])=O
S-1,2,6-TrN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@H](O)CO[N+]([O-])=O
S-1,2-DN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@H](O)CO
S-1,3,4,5,6-PN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
S-1,3,4,5-TN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](O)CO[N+]([O-])=O
S-1,3,4,6-TN	O=[N+]([O-])O[C@@H]([C@H](O[N+]([O-])=O)[C@@H](O)CO[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O

Table A. 1 (Continued): List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
S-1,3,4-TrN	O=[N+]([O-])O[C@@H]([C@H](O[N+]([O-])=O)[C@H](O)CO)[C@@H](O)CO[N+]([O-])=O
S-1,3,5,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O[N+]([O-])=O)[C@@H](O)CO[N+]([O-])=O
S-1,3,5-TrN	[O-][N+](=O)O[C@H](CO)[C@@H](O)[C@H](O[N+]([O-])=O)[C@@H](O)CO[N+]([O-])=O
S-1,3,6-TrN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@@H](O)CO[N+]([O-])=O
S-1,3-DN	O[C@@H]([C@H](O[N+](=O)[O-])[C@@H](O)CO[N+]([O-])=O)[C@H](O)CO
S-1,4,5,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](O)CO[N+]([O-])=O
S-1,4-DN	O[C@@H]([C@H](O[N+](=O)[O-])[C@H](O)CO)[C@@H](O)CO[N+]([O-])=O
S-1,5-DN	[O-][N+](=O)O[C@H](CO)[C@@H](O)[C@H](O)[C@@H](O)CO[N+]([O-])=O
S-1,6-DN	[O-][N+](=O)OC[C@H](O)[C@@H](O)[C@H](O)[C@H](O)CO[N+]([O-])=O
S-1-MN	[O-][N+](=O)OC[C@H](O)[C@@H](O)[C@H](O)[C@H](O)CO
S-2,3,4,5,6-PN	O=[N+]([O-])O[C@@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](CO)O[N+]([O-])=O
S-2,3,4,5-TN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](CO)O[N+]([O-])=O
S-2,3,4,6-TN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
S-2,3,4-TrN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@H](O)CO
S-2,3,5,6-TN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO[N+]([O-])=O)O[N+]([O-])=O
S-2,3,5-TrN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](CO)O[N+]([O-])=O
S-2,3,6-TrN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](O)CO[N+]([O-])=O
S-2,3-DN	O=[N+]([O-])O[C@@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](O)CO
S-2,4,5,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](CO)O[N+]([O-])=O
S-2,4,5-TrN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@H](CO)O[N+]([O-])=O
S-2,4,6-TrN	[O-][N+](=O)O[C@@H](CO)[C@@H](O)[C@H](O[N+]([O-])=O)[C@H](O)CO[N+]([O-])=O
S-2,4-DN	[O-][N+](=O)O[C@@H](CO)[C@@H](O)[C@H](O[N+]([O-])=O)[C@H](O)CO
S-2,5,6-TrN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@H](CO)O[N+]([O-])=O
S-2,5-DN	[O-][N+](=O)O[C@@H](CO)[C@@H](O)[C@H](O)[C@@H](CO)O[N+]([O-])=O
S-2,6-DN	[O-][N+](=O)O[C@@H](CO)[C@@H](O)[C@H](O)[C@H](O)CO[N+]([O-])=O
S-2-MN	[O-][N+](=O)O[C@@H](CO)[C@@H](O)[C@H](O)[C@H](O)CO
S-3,4,5,6-TN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](O)CO
S-3,4,5-TrN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O[N+]([O-])=O)[C@@H](O)CO
S-3,4,6-TrN	O=[N+]([O-])O[C@@H]([C@H](O[N+]([O-])=O)[C@@H](O)CO)[C@H](O)CO[N+]([O-])=O
S-3,4-DN	O=[N+]([O-])O[C@@H]([C@H](O[N+]([O-])=O)[C@@H](O)CO)[C@H](O)CO
S-3,5,6-TrN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O[N+]([O-])=O)[C@@H](O)CO
S-3,5-DN	[O-][N+](=O)O[C@H](CO)[C@@H](O)[C@H](O[N+]([O-])=O)[C@@H](O)CO
S-3,6-DN	O[C@@H]([C@H](O[N+](=O)[O-])[C@@H](O)CO)[C@H](O)CO[N+]([O-])=O
S-3-MN	O[C@@H]([C@H](O[N+](=O)[O-])[C@@H](O)CO)[C@H](O)CO
S-4,5,6-TrN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](O)CO
S-4,5-DN	O=[N+]([O-])O[C@H](CO)[C@@H](O[N+]([O-])=O)[C@H](O)[C@@H](O)CO

Table A. 1 (Continued): List of analytes and SMILES used to generate molecular descriptors

Analyte	SMILES
S-4,6-DN	O[C@@H]([C@H](O[N+](=O)[O-])[C@H](O)CO[N+]([O-])=O)[C@@H](O)CO
S-4-MN	O[C@@H]([C@H](O[N+](=O)[O-])[C@H](O)CO)[C@@H](O)CO
S-5,6-DN	O=[N+]([O-])O[C@H](CO[N+](=O)[O-])[C@@H](O)[C@H](O)[C@@H](O)CO
S-5-MN	[O-][N+](=O)O[C@H](CO)[C@@H](O)[C@H](O)[C@@H](O)CO
S-6-MN	[O-][N+](=O)OC[C@@H](O)[C@@H](O)[C@H](O)[C@@H](O)CO
Salmeterol	c1ccc(cc1)CCCCOCCCCCCNCC(c2ccc(c(c2)CO)O)O
SHN	C([C@H]([C@H]([C@@H]([C@H](CO[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-])O[N+](=O)[O-]
Sotalol	CC(C)NCC(c1ccc(cc1)NS(=O)(=O)C)O
Strychnine	c1ccc2c(c1)[C@]34CCN5[C@H]3C[C@@H]6[C@@H]7[C@@H]4N2C(=O)C[C@@H]7OCC=C6C5
Tapentadol	CC[C@@H](C1=CC(=CC=C1)O)[C@@H](C)CN(C)C
TATB	c1(c(c(c(c(c1[N+](=O)[O-])N)[N+](=O)[O-])N)[N+](=O)[O-])N
TATP	CC1(OOC(OOC(OO1)(C)C)(C)C)C
TEGDN	C(COCCO[N+](=O)[O-])OCCO[N+](=O)[O-]
Terbutaline	CC(C)(C)NCC(c1cc(cc(c1)O)O)O
Testolactone	C[C@]12CC[C@H]3[C@H]([C@@H]1CCC(=O)O2)CCC4=CC(=O)C=C[C@]34C
Tetryl	CN(c1c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
Timolol	CC(C)(C)NC[C@@H](COc1c(nsn1)N2CCOCC2)O
TMETN	CC(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-]
TNB	c1c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
TNT	Cc1c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
Tramadol	CN(C)CC1CCCCC1(c2cccc(c2)OC)O
Triamcinolone	C[C@]12C[C@@H]([C@]3([C@H]([C@@H]1C[C@H]([C@@]2(C(=O)CO)O)O)CCC4=CC(=O)C=C[C@@]43C)F)O
Triamterene	C1=CC=C(C=C1)C2=NC3=C(N=C2N)N=C(N=C3N)N
Trichlormethiazide	c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)NC(N2)C(Cl)Cl
Xipamide	Cc1cccc(c1NC(=O)c2cc(c(cc2O)Cl)S(=O)(=O)N)C

Table A. 2: Molecular descriptors and predicted retention times for ANN and MLR models using the Goryński et al. descriptors.

			Molecular descriptors				Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	R2p	BELe6	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	1,2-DNG	Training	-0.418	0.392	0.000	4.42	3.95	-0.46	3.08	-1.33	3.99	-0.42
Explosives	1,3-DNB	Training	1.619	0.391	0.000	9.04	8.89	-0.16	6.53	-2.51	7.98	-1.06
Explosives	1,3-DNG	Training	-0.418	0.366	0.000	4.02	4.06	0.04	2.83	-1.19	4.29	0.28
Explosives	2,4-DA-6-NT	Training	0.717	0.462	0.258	3.66	7.54	3.89	6.63	2.98	6.87	3.21
Explosives	2,4-DNT	Training	2.105	0.488	0.080	11.21	9.87	-1.34	8.59	-2.62	8.27	-2.94
Explosives	2,6-DA-4-NT	Training	0.717	0.461	0.220	3.16	7.38	4.22	6.48	3.33	6.66	3.50
Explosives	2,6-DNT	Training	2.105	0.471	0.098	11.04	10.15	-0.89	8.49	-2.55	8.57	-2.48
Explosives	2-Am-4,6-DNT	Training	1.358	0.460	0.275	10.67	9.25	-1.43	7.77	-2.91	8.24	-2.43
Explosives	2-NT	Training	2.211	0.513	0.000	10.29	9.49	-0.80	8.72	-1.57	7.73	-2.56
Explosives	3,5-DNA	Training	0.872	0.365	0.237	9.26	8.98	-0.28	5.88	-3.38	8.17	-1.09
Explosives	3-NT	Training	2.211	0.510	0.000	10.87	9.52	-1.35	8.69	-2.18	7.77	-3.11
Explosives	4-Am-2,6-DNT	Training	1.358	0.458	0.360	10.33	9.63	-0.71	8.06	-2.27	8.75	-1.59
Explosives	4-NT	Training	2.211	0.526	0.000	10.47	9.36	-1.11	8.85	-1.63	7.58	-2.89
Explosives	DEDPU	Training	3.632	0.798	0.799	14.81	9.95	-4.86	16.82	2.02	11.76	-3.05
Explosives	DMNB	Training	1.566	0.522	0.759	7.30	9.91	2.61	10.50	3.19	10.68	3.38
Explosives	DPA	Training	3.380	0.835	0.652	13.97	8.64	-5.33	16.21	2.25	10.00	-3.96
Explosives	EGDN	Training	0.093	0.357	0.000	5.44	5.29	-0.15	3.61	-1.83	5.39	-0.05
Explosives	ETN	Training	0.062	0.353	0.728	11.60	10.48	-1.12	6.19	-5.41	9.52	-2.08
Explosives	HMTA	Training	0.086	0.878	0.443	1.52	1.40	-0.12	10.26	8.75	1.88	0.37
Explosives	HND	Training	2.747	0.678	1.157	15.57	12.03	-3.54	15.47	-0.10	13.45	-2.12
Explosives	MHN	Training	0.031	0.390	1.025	14.55	11.90	-2.65	7.59	-6.97	10.72	-3.83
Explosives	NB	Training	1.724	0.433	0.000	8.00	8.95	0.95	7.12	-0.88	7.70	-0.29
Explosives	NG	Training	0.078	0.336	0.000	8.97	5.29	-3.68	3.38	-5.58	5.61	-3.36
Explosives	NM	Training	0.291	0.335	0.000	1.85	5.79	3.95	3.73	1.89	6.03	4.19
Explosives	NQ	Training	-0.532	0.273	0.000	1.49	3.83	2.34	1.74	0.24	5.14	3.65
Explosives	PA	Training	1.246	0.417	0.360	4.96	10.02	5.06	7.47	2.51	9.00	4.04
Explosives	PETN	Training	-0.184	0.357	0.559	12.39	8.83	-3.56	5.19	-7.20	8.03	-4.36
Explosives	PYX	Training	4.013	0.813	1.188	16.90	13.92	-2.98	19.04	2.15	14.54	-2.36

Table A. 2 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using the Goryński et al. descriptors.

			Molecular descriptors				Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	R2p	BELe6	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	R-Salt	Training	-0.325	0.477	0.228	3.54	4.66	1.12	4.90	1.36	4.49	0.95
Explosives	TATP	Training	2.867	0.624	0.761	11.99	10.47	-1.52	13.70	1.71	12.06	0.07
Explosives	TEGDN	Training	-0.169	0.499	0.371	7.61	5.26	-2.36	5.90	-1.71	5.35	-2.26
Explosives	Tetryl	Training	1.762	0.457	0.625	11.83	11.45	-0.38	9.71	-2.12	11.05	-0.77
Explosives	TMETN	Training	0.412	0.408	0.559	11.78	9.24	-2.55	6.70	-5.09	8.61	-3.18
Explosives	TNB	Training	1.513	0.362	0.360	9.81	11.33	1.52	7.39	-2.42	10.16	0.35
Explosives	TNT	Training	1.999	0.456	0.360	12.25	11.23	-1.02	9.13	-3.12	10.02	-2.23
Drugs	6-MAM	Training	1.766	0.949	1.046	4.45	9.23	4.78	16.02	11.57	7.77	3.32
Drugs	Acebutolol	Training	1.615	0.707	1.057	6.38	8.63	2.25	13.46	7.08	10.25	3.87
Drugs	Acetazolamide	Training	-1.206	0.680	0.084	2.15	-0.19	-2.34	4.83	2.68	10.09	7.94
Drugs	AICAR	Training	-2.913	0.608	0.631	1.42	0.86	-0.56	3.24	1.82	11.39	9.97
Drugs	Alprenolol	Training	2.640	0.717	0.832	11.41	8.61	-2.81	14.47	3.06	7.49	-3.92
Drugs	Altizide	Training	1.486	0.973	0.731	6.87	5.74	-1.13	14.62	7.75	9.30	2.43
Drugs	Andarine	Training	2.315	0.695	0.978	12.40	9.20	-3.20	14.25	1.85	10.33	-2.06
Drugs	Atenolol	Training	0.669	0.701	0.799	2.73	4.69	1.96	10.85	8.12	10.66	7.93
Drugs	Beclomethasone	Training	1.905	0.903	1.162	11.52	10.65	-0.87	16.24	4.72	9.23	-2.28
Drugs	Benzoylecgonine	Training	1.868	0.909	0.880	4.05	6.96	2.92	15.20	11.15	7.49	3.44
Drugs	Benzthiazide	Training	2.498	0.956	1.008	9.39	9.82	0.42	17.19	7.80	22.36	12.97
Drugs	Beta\dexameth.	Training	1.642	0.876	1.161	10.89	10.08	-0.81	15.52	4.63	9.03	-1.87
Drugs	Betaxolol	Training	2.577	0.793	1.019	11.60	9.34	-2.26	15.79	4.19	3.88	-7.72
Drugs	Bisoprolol	Training	2.031	0.791	1.017	9.59	8.34	-1.25	14.83	5.24	5.51	-4.09
Drugs	Budesonide	Training	2.132	0.925	1.255	15.29	12.69	-2.60	17.18	1.89	9.95	-5.33
Drugs	Bumetanid	Training	2.960	0.850	0.942	6.25	9.27	3.01	16.71	10.46	10.61	4.36
Drugs	Butizide	Training	1.577	0.877	0.761	6.73	5.27	-1.46	13.95	7.23	7.57	0.85
Drugs	Canrenone	Training	3.379	0.967	1.243	14.71	14.43	-0.28	19.66	4.95	9.86	-4.85
Drugs	Carphedon	Training	0.254	0.801	0.664	5.16	2.37	-2.79	10.61	5.45	4.36	-0.81
Drugs	Celiprolol	Training	1.597	0.720	1.195	7.66	10.39	2.73	14.06	6.40	3.30	-4.36
Drugs	Chlorexolone	Training	2.348	0.925	0.895	9.74	8.04	-1.70	16.22	6.48	7.14	-2.60
Drugs	Chlorothiazide	Training	0.450	0.767	0.443	2.08	2.17	0.09	9.81	7.73	7.55	5.48

Table A. 2 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using the Goryński et al. descriptors.

			Molecular descriptors				Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	R2p	BELe6	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Chlorthalidone	Training	1.507	0.932	0.702	4.91	5.13	0.22	14.15	9.24	14.42	9.51
Drugs	Clobenzorex	Training	4.314	0.963	0.799	12.70	11.21	-1.49	19.58	6.88	11.19	-1.51
Drugs	Clobetasol	Training	2.845	0.928	1.146	12.94	11.90	-1.04	18.02	5.07	10.69	-2.25
Drugs	Clopamide	Training	2.318	0.885	0.995	6.01	8.77	2.76	16.15	10.14	12.67	6.67
Drugs	Cyclopenthiazide	Training	2.118	1.034	1.047	9.61	10.89	1.28	17.44	7.83	4.78	-4.83
Drugs	Deflazacort	Training	1.603	0.856	1.290	13.62	12.14	-1.48	15.74	2.12	9.91	-3.70
Drugs	Desonide	Training	0.962	0.896	1.237	12.50	10.92	-1.58	14.84	2.34	7.90	-4.60
Drugs	Diacetolol	Training	0.492	0.650	0.873	3.62	5.76	2.14	10.33	6.71	-0.40	-4.01
Drugs	Dichlorphenamide	Training	0.815	0.768	0.105	4.19	3.15	-1.04	9.20	5.01	9.44	5.25
Drugs	Efaproxiral	Training	3.858	0.831	0.979	7.28	11.43	4.15	18.19	10.90	11.39	4.11
Drugs	Etamivan	Training	1.658	0.637	0.701	5.75	7.25	1.49	11.55	5.80	9.20	3.45
Drugs	Famprofazone	Training	4.847	0.901	1.149	17.86	15.02	-2.84	21.17	3.31	14.93	-2.92
Drugs	Fenbutrazate	Training	4.455	0.959	1.152	17.61	14.66	-2.95	21.07	3.46	13.51	-4.10
Drugs	Fencamfamin	Training	3.086	1.014	0.799	10.22	9.31	-0.92	17.98	7.76	8.20	-2.02
Drugs	Fenetylline	Training	1.785	0.889	0.996	9.46	7.98	-1.48	15.29	5.83	8.22	-1.24
Drugs	Fenfluramine	Training	3.358	0.721	0.521	9.29	9.69	0.39	14.59	5.30	10.53	1.24
Drugs	Fludrocortisone	Training	1.346	0.885	1.154	9.59	9.70	0.11	15.08	5.49	8.30	-1.29
Drugs	Flumethasone	Training	1.480	0.861	1.159	11.13	9.74	-1.39	15.10	3.96	8.87	-2.26
Drugs	Fluticasone prop.	Training	3.947	0.887	1.241	16.21	14.65	-1.56	19.84	3.63	13.86	-2.35
Drugs	FPCAM	Training	2.999	0.877	1.231	10.79	13.02	2.23	18.10	7.31	12.06	1.27
Drugs	Fulvestrant	Training	8.412	1.022	1.476	18.08	22.97	4.89	29.60	11.52	10.93	-7.15
Drugs	Furosemide	Training	1.714	0.715	0.753	3.36	6.23	2.86	12.59	9.23	15.12	11.75
Drugs	Gestrinone	Training	4.132	0.981	1.160	14.00	14.43	0.43	20.77	6.77	0.21	-13.79
Drugs	Hydrochloroth.	Training	0.038	0.764	0.478	2.13	1.49	-0.64	9.21	7.08	8.00	5.88
Drugs	Hydroflumeth.	Training	0.316	0.678	0.555	2.63	3.22	0.59	9.13	6.50	7.75	5.12
Drugs	Isometheptene	Training	2.405	0.600	0.490	6.99	9.53	2.54	11.69	4.70	9.89	2.90
Drugs	Mefenorex	Training	2.793	0.857	0.772	8.50	7.70	-0.80	15.87	7.37	9.28	0.78
Drugs	Methylprednisolone	Training	1.446	0.874	1.161	11.44	9.84	-1.61	15.17	3.73	8.67	-2.77
Drugs	Ostarine	Training	2.315	0.695	0.978	13.34	9.20	-4.14	14.25	0.91	7.75	-5.59

Table A. 2 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using the Goryński et al. descriptors.

			Molecular descriptors				Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	R2p	BELe6	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Oxycodone	Training	1.029	0.892	0.982	4.60	6.81	2.21	13.98	9.38	6.63	2.02
Drugs	Oxymorphone	Training	0.778	0.952	0.972	2.94	7.14	4.20	14.10	11.16	5.39	2.45
Drugs	Pentazocine	Training	4.348	0.851	0.992	8.59	12.51	3.92	19.26	10.67	13.64	5.05
Drugs	Pentetrazol	Training	0.918	0.826	0.000	2.86	3.40	0.54	9.55	6.69	1.59	-1.27
Drugs	Phendimetrazine	Training	1.998	0.744	0.799	6.32	6.72	0.39	13.52	7.20	9.19	2.87
Drugs	Phenmetrazine	Training	1.462	0.773	0.472	5.19	4.30	-0.89	11.69	6.50	5.95	0.75
Drugs	Prenylamine	Training	5.711	1.025	0.952	15.88	14.97	-0.91	23.11	7.23	14.07	-1.81
Drugs	Probenecid	Training	2.821	0.791	0.799	5.30	8.13	2.83	15.38	10.08	9.86	4.56
Drugs	Propylhexedrine	Training	2.739	0.753	0.545	8.76	7.66	-1.11	13.94	5.17	9.09	0.33
Drugs	Salmeterol	Training	4.226	0.833	1.257	15.58	15.18	-0.40	19.85	4.27	11.00	-4.58
Drugs	Sotalol	Training	0.967	0.728	0.761	2.91	4.63	1.72	11.48	8.57	5.16	2.25
Drugs	Tapentadol	Training	3.441	0.705	0.809	7.33	10.53	3.20	15.64	8.30	12.52	5.18
Drugs	Terbutaline	Training	1.254	0.702	0.761	2.83	5.50	2.67	11.71	8.88	9.95	7.12
Drugs	Testolactone	Training	3.123	0.928	0.985	9.17	10.31	1.14	17.90	8.73	12.61	3.44
Drugs	Timolol	Training	1.128	0.800	0.924	7.95	5.83	-2.12	13.04	5.10	6.61	-1.33
Drugs	Tramadol	Training	2.701	0.785	1.009	7.12	9.50	2.37	15.89	8.76	11.28	4.16
Drugs	Triamcinolone	Training	0.356	0.858	1.146	7.24	8.41	1.18	13.11	5.87	6.64	-0.60
Drugs	Triamterene	Training	1.017	0.711	0.785	6.56	5.07	-1.49	11.49	4.93	9.74	3.19
Drugs	Xipamide	Training	2.812	0.841	0.939	5.01	8.94	3.93	16.36	11.35	10.32	5.31
Explosives	1,2-DNB	Select	1.619	0.478	0.000	7.96	8.35	0.38	7.37	-0.59	6.98	-0.98
Explosives	3,4-DNT	Select	2.105	0.491	0.074	10.13	9.81	-0.33	8.60	-1.54	8.20	-1.93
Explosives	HMDD	Select	1.738	0.699	0.149	4.43	6.03	1.60	10.26	5.83	5.51	1.08
Explosives	PGDN	Select	0.471	0.400	0.000	7.50	6.04	-1.46	4.67	-2.83	5.64	-1.87
Explosives	RDX	Select	0.623	0.474	0.563	5.95	8.33	2.38	7.71	1.76	8.28	2.32
Explosives	TATB	Select	-0.727	0.420	0.699	9.40	7.62	-1.79	5.39	-4.01	7.04	-2.37
Drugs	Amiloride	Select	-0.077	0.490	0.220	3.63	5.00	1.36	5.41	1.78	9.20	5.57
Drugs	Aminoglutethimide	Select	1.285	0.748	0.799	4.99	5.26	0.27	12.35	7.36	9.15	4.16
Drugs	Bendroflumethiazide	Select	2.179	0.923	0.802	8.50	6.92	-1.59	15.57	7.07	8.30	-0.21
Drugs	Bumetanide	Select	2.960	0.850	0.942	6.28	9.27	2.98	16.71	10.43	9.29	3.01

Table A. 2 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using the Goryński et al. descriptors.

			Molecular descriptors				Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	R2p	BELe6	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Bunolol	Select	2.358	0.806	0.795	7.59	7.02	-0.58	14.72	7.13	2.66	-4.94
Drugs	Danazol	Select	4.875	0.937	1.123	16.65	14.89	-1.76	21.47	4.82	12.84	-3.81
Drugs	Desacetyl deflazacort	Select	1.224	0.887	1.217	10.85	10.69	-0.16	15.12	4.27	8.40	-2.45
Drugs	Esmolol	Select	1.978	0.701	0.837	8.36	7.41	-0.95	13.21	4.85	5.28	-3.08
Drugs	Etacr. acid	Select	4.152	0.815	0.763	7.31	10.91	3.60	17.74	10.43	9.71	2.41
Drugs	Labetalol	Select	2.330	0.840	1.038	9.84	9.11	-0.73	15.89	6.05	10.45	0.61
Drugs	Metoprolol	Select	1.757	0.702	0.799	7.27	6.72	-0.55	12.71	5.44	8.91	1.64
Drugs	Norfenfluramine	Select	2.577	0.665	0.348	8.60	8.67	0.07	12.09	3.49	8.67	0.07
Drugs	Pethidine	Select	2.446	0.780	0.799	8.41	7.37	-1.04	14.63	6.22	9.65	1.24
Drugs	Propranolol	Select	2.540	0.761	0.765	11.61	7.58	-4.03	14.48	2.88	3.35	-8.26
Drugs	Strychnine	Select	1.146	1.073	1.050	8.14	10.73	2.59	16.18	8.03	5.15	-2.99
Drugs	Trichlormethiazide	Select	1.364	1.003	0.515	4.92	4.77	-0.15	13.91	8.99	8.33	3.41
Explosives	1-MNG	Test	-0.913	0.387	0.000	1.85	2.84	0.99	2.19	0.34	3.08	1.23
Explosives	DADP	Test	1.911	0.583	0.000	8.43	7.94	-0.50	8.89	0.45	6.34	-2.09
Explosives	DMDPU	Test	2.935	0.712	0.799	12.44	9.20	-3.23	14.81	2.37	11.39	-1.05
Explosives	HMTD	Test	2.558	0.671	0.619	2.80	8.54	5.75	13.11	10.31	10.10	7.31
Explosives	HMX	Test	0.831	0.563	0.627	4.64	6.89	2.25	9.16	4.52	8.02	3.39
Explosives	SHN	Test	0.031	0.391	1.025	14.55	11.88	-2.67	7.60	-6.96	10.71	-3.85
Drugs	Anastrazole	Test	2.966	0.841	0.761	8.02	8.05	0.03	15.97	7.95	7.16	-0.86
Drugs	Bambuterol	Test	1.727	0.716	0.880	8.41	7.00	-1.41	13.09	4.68	10.94	2.53
Drugs	Bupropion	Test	3.227	0.816	0.761	9.54	8.74	-0.81	16.17	6.63	10.54	1.00
Drugs	Carteolol	Test	1.283	0.781	0.763	4.17	4.79	0.62	12.53	8.36	12.37	8.20
Drugs	Carvedilol	Test	4.015	0.933	1.091	13.30	13.07	-0.23	19.85	6.55	8.71	-4.59
Drugs	Epitizide	Test	1.876	0.926	0.658	6.68	5.47	-1.20	14.56	7.88	10.66	3.98
Drugs	Formoterol	Test	1.929	0.811	1.051	7.71	8.58	0.87	14.98	7.27	7.09	-0.62
Drugs	Mefruside	Test	1.528	0.926	0.925	9.61	7.09	-2.52	14.95	5.34	6.88	-2.73
Drugs	Nadoxolol	Test	1.369	0.631	0.629	9.26	6.53	-2.73	10.74	1.48	7.48	-1.78
Drugs	Nikethamide	Test	0.791	0.697	0.486	3.61	3.75	0.14	9.87	6.25	5.59	1.98
Drugs	Oxprenolol	Test	2.232	0.674	0.827	9.46	8.32	-1.14	13.35	3.89	6.05	-3.41

Table A. 2 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using the Goryński et al. descriptors.

			Molecular descriptors				Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	R2p	BELe6	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Pemoline	Test	1.336	0.714	0.212	3.15	4.77	1.63	9.95	6.81	4.91	1.76
Drugs	Pindolol	Test	1.926	0.738	0.830	4.86	6.81	1.95	13.46	8.59	8.90	4.04
Drugs	Piretanide	Test	2.446	0.871	0.938	5.12	8.31	3.19	16.02	10.91	9.39	4.27
Drugs	Polythiazide	Test	2.412	0.881	0.884	8.84	7.78	-1.06	15.86	7.03	11.84	3.00
Drugs	Prolintane	Test	4.043	0.870	0.799	9.25	10.57	1.32	18.22	8.97	11.73	2.48

Table A. 3: Molecular descriptors and predicted retention time for best ANN model using the Barron & McEneff descriptors.

			Molecular descriptors															Average	Best ANN	
Dataset	Analyte	ANN subset	nDB	nC	nO	nR03	nR04	nR05	nR06	nR07	nR08	nR09	nBnz	Ui	Hy	MLOGP	ALOGP	LogD (pH=7)	t_R^M (min)	t_R^P (min)	Δt_R (min)
Explosives	1,2-DNB	Training	4	6	4	0	0	0	1	0	0	0	1	3.459	-0.484	1.886	1.619	1.60	7.96	8.35	0.39
Explosives	1,2-DNG	Training	4	3	7	0	0	0	0	0	0	0	0	2.322	0.321	-0.164	-0.418	0.86	4.42	5.29	0.88
Explosives	1,3-DNB	Training	4	6	4	0	0	0	1	0	0	0	1	3.459	-0.484	1.886	1.619	1.55	9.04	8.30	-0.75
Explosives	1,3-DNG	Training	4	3	7	0	0	0	0	0	0	0	0	2.322	0.321	-0.164	-0.418	0.97	4.02	5.43	1.42
Explosives	1-MNG	Training	2	3	5	0	0	0	0	0	0	0	0	1.585	1.161	-0.827	-0.913	-0.43	1.85	2.42	0.57
Explosives	2,4-DNT	Training	4	7	4	0	0	0	1	0	0	0	1	3.459	-0.523	2.239	2.105	2.10	11.21	9.48	-1.73
Explosives	2,6-DA-4-NT	Training	2	7	2	0	0	0	1	0	0	0	1	3.170	2.617	1.181	0.717	0.92	3.16	5.88	2.72
Explosives	2,6-DNT	Training	4	7	4	0	0	0	1	0	0	0	1	3.459	-0.523	2.239	2.105	2.10	11.04	9.48	-1.56
Explosives	2-NT	Training	2	7	2	0	0	0	1	0	0	0	1	3.170	-0.672	2.241	2.211	2.45	10.29	9.25	-1.04
Explosives	3,4-DNT	Training	4	7	4	0	0	0	1	0	0	0	1	3.459	-0.523	2.239	2.105	2.15	10.13	9.53	-0.60
Explosives	3,5-DNA	Training	4	6	4	0	0	0	1	0	0	0	1	3.459	0.829	1.395	0.872	1.63	9.26	7.57	-1.69
Explosives	4-Am-2,6-DNT	Training	4	7	4	0	0	0	1	0	0	0	1	3.459	0.756	1.748	1.358	1.88	10.33	8.44	-1.90
Explosives	4-NT	Training	2	7	2	0	0	0	1	0	0	0	1	3.170	-0.672	2.241	2.211	2.45	10.47	9.25	-1.22
Explosives	DADP	Training	0	6	4	0	0	0	1	0	0	0	0	0.000	-0.576	1.107	1.911	1.23	8.43	7.04	-1.40
Explosives	DMDPU	Training	1	15	1	0	0	0	2	0	0	0	2	3.807	-0.818	3.202	2.935	2.55	12.44	12.31	-0.13
Explosives	DPA	Training	0	12	0	0	0	0	2	0	0	0	2	3.700	-0.352	3.395	3.380	3.12	13.97	11.98	-1.99
Explosives	EGDN	Training	4	2	6	0	0	0	0	0	0	0	0	2.322	-0.192	0.056	0.093	1.51	5.44	6.13	0.69
Explosives	ETN	Training	8	4	12	0	0	0	0	0	0	0	0	3.170	-0.197	1.125	0.062	3.21	11.60	11.41	-0.19
Explosives	HMDD	Training	0	6	4	0	0	0	0	0	2	0	0	0.000	-0.484	0.359	1.738	0.00	4.43	2.75	-1.69
Explosives	HMTA	Training	0	6	0	0	0	0	4	0	3	0	0	0.000	-0.576	1.107	0.086	0.98	1.52	1.70	0.19
Explosives	HMTD	Training	0	6	6	0	0	0	0	0	0	0	0	0.000	-0.418	0.476	2.558	0.13	2.80	5.44	2.64
Explosives	HMX	Training	8	4	8	0	0	0	0	0	1	0	0	3.170	-0.197	2.357	0.831	-1.71	4.64	5.20	0.57
Explosives	HND	Training	12	12	12	0	0	0	2	0	0	0	2	4.644	0.023	3.141	2.747	4.44	15.57	14.64	-0.93
Explosives	NG	Training	6	3	9	0	0	0	0	0	0	0	0	2.807	-0.195	0.575	0.078	2.41	8.97	8.86	-0.11
Explosives	NM	Training	2	1	2	0	0	0	0	0	0	0	0	1.585	-0.215	-0.433	0.291	0.10	1.85	3.19	1.34
Explosives	NQ	Training	3	1	2	0	0	0	0	0	0	0	0	2.000	3.831	-0.915	-0.532	-2.56	1.49	0.88	-0.61
Explosives	PA	Training	6	6	7	0	0	0	1	0	0	0	1	3.700	0.138	1.543	1.246	-1.40	4.96	4.83	-0.13

Table A. 3 (Continued): Molecular descriptors and predicted retention time for best ANN model using the Barron & McEneff descriptors.

			Molecular descriptors															Average	Best ANN	
Dataset	Analyte	ANN subset	nDB	nC	nO	nR03	nR04	nR05	nR06	nR07	nR08	nR09	nBnz	Ui	Hy	MLOGP	ALOGP	LogD (pH=7)	t_R^M (min)	t_R^P (min)	Δt_R (min)
Explosives	PETN	Training	8	5	12	0	0	0	0	0	0	0	0	3.170	-0.235	1.535	-0.184	3.64	12.39	12.64	0.25
Explosives	PGDN	Training	4	3	6	0	0	0	0	0	0	0	0	2.322	-0.263	0.575	0.471	1.87	7.50	7.45	-0.05
Explosives	PYX	Training	16	17	16	0	0	0	3	0	0	0	2	5.129	0.488	4.266	4.013	7.04	16.90	18.11	1.21
Explosives	R-Salt	Training	3	3	3	0	0	0	1	0	0	0	0	2.000	-0.242	-0.635	-0.325	-1.18	3.54	1.30	-2.24
Explosives	SHN	Training	12	6	18	0	0	0	0	0	0	0	0	3.700	-0.198	2.301	0.031	5.16	14.55	15.98	1.43
Explosives	TATB	Training	6	6	6	0	0	0	1	0	0	0	1	3.700	4.331	0.102	-0.727	2.93	9.40	8.33	-1.08
Explosives	TEGDN	Training	4	6	8	0	0	0	0	0	0	0	0	2.322	-0.367	0.346	-0.169	1.46	7.61	7.31	-0.30
Drugs	Acebutolol	Training	2	18	4	0	0	0	1	0	0	0	1	3.170	0.998	1.589	1.615	-0.73	6.38	6.93	0.55
Drugs	Acetazolamide	Training	3	4	3	0	0	1	0	0	0	0	0	3.170	1.837	-1.583	-1.206	-0.58	2.15	3.06	0.91
Drugs	Alprenolol	Training	1	15	2	0	0	0	1	0	0	0	1	3.000	0.320	2.370	2.640	0.34	11.41	8.38	-3.03
Drugs	Altizide	Training	5	11	4	0	0	0	2	0	0	0	1	3.585	2.094	0.568	1.486	1.08	6.87	5.70	-1.17
Drugs	Amiloride	Training	2	6	1	0	0	0	1	0	0	0	0	3.170	6.789	0.157	-0.077	-0.92	3.63	3.35	-0.29
Drugs	Aminoglutethimide	Training	2	13	2	0	0	0	2	0	0	0	1	3.170	1.193	1.377	1.285	1.27	4.99	6.76	1.77
Drugs	Anastrazole	Training	0	17	0	0	0	1	1	0	0	0	1	3.807	-0.762	2.795	2.966	2.68	8.02	11.29	3.27
Drugs	Atenolol	Training	1	14	3	0	0	0	1	0	0	0	1	3.000	1.986	0.925	0.669	-2.19	2.73	3.32	0.59
Drugs	Beclomethasone	Training	4	22	5	0	0	1	3	0	0	1	0	2.322	0.882	1.962	1.905	2.25	11.52	11.30	-0.22
Drugs	Benzoylecgonine	Training	2	16	4	0	0	1	2	1	0	0	1	3.170	-0.291	2.170	1.868	-0.21	4.05	3.46	-0.59
Drugs	Benzthiazide	Training	5	15	4	0	0	0	3	0	0	0	2	4.170	1.126	1.793	2.498	0.99	9.39	8.11	-1.28
Drugs	Beta\dexameth.	Training	4	22	5	0	0	1	3	0	0	1	0	2.322	0.882	1.855	1.642	1.92	10.89	10.88	-0.01
Drugs	Betaxolol	Training	0	18	3	1	0	0	1	0	0	0	1	2.807	0.259	1.991	2.577	0.43	11.60	11.24	-0.36
Drugs	Bisoprolol	Training	0	18	4	0	0	0	1	0	0	0	1	2.807	0.278	1.595	2.031	-0.22	9.59	7.60	-1.99
Drugs	Budesonide	Training	4	25	6	0	0	2	3	0	1	1	0	2.322	0.161	2.014	2.132	3.02	15.29	13.36	-1.93
Drugs	Bupropion	Training	1	13	1	0	0	0	1	0	0	0	1	3.000	-0.291	3.218	3.227	2.69	9.54	11.13	1.59
Drugs	Canrenone	Training	4	22	3	0	0	2	3	0	0	1	0	2.322	-0.869	4.154	3.379	2.50	14.71	15.58	0.87
Drugs	Celiprolol	Training	2	20	4	0	0	0	1	0	0	0	1	3.170	0.945	1.696	1.597	-0.41	7.66	7.81	0.15
Drugs	Chlorexolone	Training	3	14	3	0	0	1	2	0	0	1	1	3.322	0.425	1.968	2.348	1.49	9.74	9.22	-0.52
Drugs	Chlorothiazide	Training	5	7	4	0	0	0	2	0	0	0	1	3.585	1.539	0.187	0.450	-0.10	2.08	3.47	1.39
Drugs	Clobenzorex	Training	0	16	0	0	0	0	2	0	0	0	2	3.700	-0.389	4.411	4.314	2.89	12.70	13.56	0.86

Table A. 3 (Continued): Molecular descriptors and predicted retention time for best ANN model using the Barron & McEneff descriptors.

			Molecular descriptors															Average	Best ANN	
Dataset	Analyte	ANN subset	nDB	nC	nO	nR03	nR04	nR05	nR06	nR07	nR08	nR09	nBnz	Ui	Hy	MLOGP	ALOGP	LogD (pH=7)	t_R^M (min)	t_R^P (min)	Δt_R (min)
Drugs	Clobetasol	Training	4	22	4	0	0	1	3	0	0	1	0	2.322	0.212	2.855	2.845	2.74	12.94	12.67	-0.27
Drugs	Clopamide	Training	3	14	3	0	0	0	2	0	0	0	1	3.322	1.159	2.008	2.318	1.75	6.01	8.37	2.37
Drugs	Cyclopenthiazide	Training	4	13	4	0	0	1	2	0	0	0	1	3.459	1.993	1.181	2.118	1.51	9.61	7.71	-1.90
Drugs	Deflazacort	Training	6	25	6	0	0	2	3	0	1	1	0	2.807	-0.372	2.305	1.603	2.43	13.62	13.79	0.17
Drugs	Desacetyl deflazacort	Training	5	23	5	0	0	2	3	0	1	1	0	2.585	0.194	1.916	1.224	1.84	10.85	12.30	1.45
Drugs	Desonide	Training	4	24	6	0	0	2	3	0	1	1	0	2.322	0.177	1.807	0.962	2.79	12.50	12.82	0.32
Drugs	Diacetolol	Training	2	16	4	0	0	0	1	0	0	0	1	3.170	1.069	1.108	0.492	-1.40	3.62	5.19	1.58
Drugs	Dichlorphenamide	Training	4	6	4	0	0	0	1	0	0	0	1	3.459	2.504	0.198	0.815	0.89	4.19	5.03	0.84
Drugs	Efaproxiral	Training	2	20	4	0	0	0	2	0	0	0	2	3.907	0.233	3.125	3.858	0.38	7.28	11.10	3.82
Drugs	Epitizide	Training	4	10	4	0	0	0	2	0	0	0	1	3.459	2.107	0.783	1.876	1.01	6.68	5.51	-1.17
Drugs	Esmolol	Training	1	16	4	0	0	0	1	0	0	0	1	3.000	0.331	1.836	1.978	-0.42	8.36	7.34	-1.02
Drugs	Etamivan	Training	1	12	3	0	0	0	1	0	0	0	1	3.000	-0.229	1.713	1.658	1.44	5.75	8.34	2.59
Drugs	Famprofazone	Training	2	24	1	0	0	1	2	0	0	0	2	3.907	-0.848	4.389	4.847	1.18	17.86	13.38	-4.48
Drugs	Fencamfamin	Training	0	15	0	0	0	2	2	0	0	0	1	2.807	-0.413	3.517	3.086	0.39	10.22	9.46	-0.76
Drugs	Fludrocortisone	Training	3	21	5	0	0	1	3	0	0	1	0	2.000	0.908	1.728	1.346	1.54	9.59	9.98	0.39
Drugs	Flumethasone	Training	4	22	5	0	0	1	3	0	0	1	0	2.322	0.891	1.962	1.480	1.74	11.13	10.84	-0.29
Drugs	Fulvestrant	Training	1	32	3	0	0	1	3	0	0	1	1	3.000	0.113	6.073	8.412	7.56	18.08	18.21	0.13
Drugs	Gestrinone	Training	4	21	2	0	0	1	3	0	0	1	0	2.585	-0.455	3.873	4.132	3.26	14.00	13.65	-0.35
Drugs	Hydrochloroth.	Training	4	7	4	0	0	0	2	0	0	0	1	3.459	2.409	-0.547	0.038	0.00	2.13	2.31	0.19
Drugs	Hydroflumeth.	Training	4	8	4	0	0	0	2	0	0	0	1	3.459	2.272	-0.077	0.316	0.23	2.63	3.35	0.72
Drugs	Labetalol	Training	1	19	3	0	0	0	2	0	0	0	2	3.807	2.578	2.674	2.330	0.53	9.84	9.01	-0.83
Drugs	Methylprednisolone	Training	4	22	5	0	0	1	3	0	0	1	0	2.322	0.872	1.747	1.446	1.97	11.44	10.81	-0.63
Drugs	Metoprolol	Training	0	15	3	0	0	0	1	0	0	0	1	2.807	0.341	1.653	1.757	-0.59	7.27	6.62	-0.65
Drugs	Nadoxolol	Training	1	14	3	0	0	0	2	0	0	0	2	3.700	1.986	1.974	1.369	1.63	9.26	8.61	-0.65
Drugs	Nikethamide	Training	1	10	1	0	0	0	1	0	0	0	0	3.000	-0.748	1.020	0.791	0.54	3.61	5.26	1.65
Drugs	Ostarine	Training	4	19	6	0	0	0	2	0	0	0	2	4.087	1.004	2.447	2.315	3.26	13.34	12.21	-1.13
Drugs	Oxprenolol	Training	1	15	3	0	0	0	1	0	0	0	1	3.000	0.341	1.831	2.232	-0.15	9.46	7.30	-2.16
Drugs	Oxycodone	Training	1	18	4	0	0	1	4	0	1	3	1	3.000	-0.326	1.381	1.029	0.15	4.60	3.79	-0.81

Table A. 3 (Continued): Molecular descriptors and predicted retention time for best ANN model using the Barron & McEneff descriptors.

			Molecular descriptors															Average	Best ANN	
Dataset	Analyte	ANN subset	nDB	nC	nO	nR03	nR04	nR05	nR06	nR07	nR08	nR09	nBnz	Ui	Hy	MLOGP	ALOGP	LogD (pH=7)	t_R^M (min)	t_R^P (min)	Δt_R (min)
Drugs	Pemoline	Training	2	9	2	0	0	1	1	0	0	0	1	3.170	0.605	1.179	1.336	0.50	3.15	6.02	2.88
Drugs	Pentazocine	Training	1	19	1	0	0	0	3	0	1	0	1	3.000	-0.432	3.777	4.348	1.89	8.59	8.81	0.21
Drugs	Pentetrazol	Training	0	6	0	0	0	1	0	1	0	0	0	2.585	-0.576	1.666	0.918	0.22	2.86	2.57	-0.29
Drugs	Pethidine	Training	1	15	2	0	0	0	2	0	0	0	1	3.000	-0.818	2.522	2.446	1.56	8.41	9.55	1.14
Drugs	Pindolol	Training	0	14	2	0	0	1	1	0	0	1	1	3.459	1.142	1.306	1.926	-0.65	4.86	5.47	0.61
Drugs	Piretanide	Training	3	17	5	0	0	1	2	0	0	0	2	4.000	1.045	1.371	2.446	0.07	5.12	6.00	0.89
Drugs	Polythiazide	Training	4	11	4	0	0	0	2	0	0	0	1	3.459	1.282	1.051	2.412	1.59	8.84	6.66	-2.18
Drugs	Prenylamine	Training	0	24	0	0	0	0	3	0	0	0	3	4.248	-0.514	5.475	5.711	2.60	15.88	17.69	1.81
Drugs	Probenecid	Training	3	13	4	0	0	0	1	0	0	0	1	3.322	-0.198	1.979	2.821	0.13	5.30	7.78	2.48
Drugs	Prolintane	Training	0	15	0	0	0	1	1	0	0	0	1	2.807	-0.917	3.517	4.043	1.36	9.25	11.15	1.90
Drugs	Propranolol	Training	0	16	2	0	0	0	2	0	0	0	2	3.585	0.290	2.534	2.540	0.81	11.61	9.93	-1.67
Drugs	Propylhexedrine	Training	0	10	0	0	0	0	1	0	0	0	0	0.000	-0.294	2.502	2.739	-0.30	8.76	8.10	-0.67
Drugs	Salmeterol	Training	0	25	4	0	0	0	2	0	0	0	2	3.700	1.531	2.872	4.226	1.62	15.58	12.05	-3.53
Drugs	Sotalol	Training	2	12	3	0	0	0	1	0	0	0	1	3.170	1.251	0.709	0.967	-1.97	2.91	3.09	0.18
Drugs	Strychnine	Training	2	21	2	0	0	2	4	1	1	6	1	3.170	-0.830	2.904	1.146	0.21	8.14	8.88	0.74
Drugs	Tapentadol	Training	0	14	1	0	0	0	1	0	0	0	1	2.807	-0.352	3.040	3.441	0.28	7.33	9.12	1.79
Drugs	Terbutaline	Training	0	12	3	0	0	0	1	0	0	0	1	2.807	2.137	1.129	1.254	-1.94	2.83	3.34	0.51
Drugs	Testolactone	Training	4	19	3	0	0	0	4	0	0	0	0	2.322	-0.851	3.485	3.123	3.07	9.17	10.34	1.17
Drugs	Timolol	Training	0	13	3	0	0	1	1	0	0	0	0	2.585	0.472	1.223	1.128	-0.71	7.95	5.31	-2.64
Drugs	Triamcinolone	Training	4	21	6	0	0	1	3	0	0	1	0	2.322	1.662	0.866	0.356	0.92	7.24	8.41	1.18
Drugs	Triamterene	Training	0	12	0	0	0	0	3	0	0	0	1	4.170	3.956	2.703	1.017	1.10	6.56	5.31	-1.25
Drugs	Xipamide	Training	3	15	4	0	0	0	2	0	0	0	2	4.000	1.908	2.384	2.812	1.15	5.01	8.98	3.97
Explosives	2-Am-4,6-DNT	Select	4	7	4	0	0	0	1	0	0	0	1	3.459	0.756	1.748	1.358	1.88	10.67	8.44	-2.23
Explosives	DEDPU	Select	1	17	1	0	0	0	2	0	0	0	2	3.807	-0.836	3.695	3.632	3.26	14.81	13.62	-1.19
Explosives	NB	Select	2	6	2	0	0	0	1	0	0	0	1	3.170	-0.636	1.887	1.724	1.81	8.00	8.02	0.03
Explosives	RDX	Select	6	3	6	0	0	0	1	0	0	0	0	2.807	-0.195	1.401	0.623	-1.20	5.95	4.84	-1.11
Explosives	TATP	Select	0	9	6	0	0	0	0	0	0	1	0	0.000	-0.586	1.769	2.867	2.16	11.99	10.32	-1.67
Explosives	Tetryl	Select	8	7	8	0	0	0	1	0	0	0	1	3.907	-0.344	2.654	1.762	1.25	11.83	10.42	-1.41

Table A. 3 (Continued): Molecular descriptors and predicted retention time for best ANN model using the Barron & McEneff descriptors.

			Molecular descriptors															Average	Best ANN	
Dataset	Analyte	ANN subset	nDB	nC	nO	nR03	nR04	nR05	nR06	nR07	nR08	nR09	nBnz	Ui	Hy	MLOGP	ALOGP	LogD (pH=7)	t_R^M (min)	t_R^P (min)	Δt_R (min)
Explosives	TNB	Select	6	6	6	0	0	0	1	0	0	0	1	3.700	-0.391	2.005	1.513	1.36	9.81	8.87	-0.95
Drugs	Bambuterol	Select	2	18	5	0	0	0	1	0	0	0	1	3.170	0.327	1.032	1.727	-0.93	8.41	6.17	-2.24
Drugs	Bendroflumethiazide	Select	4	15	4	0	0	0	3	0	0	0	2	4.087	1.881	1.582	2.179	1.48	8.50	7.92	-0.58
Drugs	Bumetanide	Select	3	17	5	0	0	0	2	0	0	0	2	4.000	1.815	1.763	2.960	0.21	6.25	7.79	1.54
Drugs	Bunolol	Select	1	17	3	0	0	0	2	0	0	0	1	3.000	0.284	2.323	2.358	0.15	7.59	8.08	0.48
Drugs	Carphedon	Select	2	12	2	0	0	1	1	0	0	0	1	3.170	0.451	0.837	0.254	0.31	5.16	6.09	0.93
Drugs	Carvedilol	Select	0	24	4	0	0	1	3	0	0	2	3	4.459	0.834	2.195	4.015	2.67	13.30	11.83	-1.47
Drugs	Danazol	Select	1	22	2	0	0	2	3	0	0	2	0	3.000	-0.435	4.154	4.875	4.33	16.65	16.14	-0.51
Drugs	Furosemide	Select	3	12	5	0	0	1	1	0	0	0	1	3.907	2.062	0.434	1.714	-0.74	3.36	3.66	0.30
Drugs	Isometheptene	Select	1	9	0	0	0	0	0	0	0	0	0	1.000	-0.257	2.446	2.405	-0.03	6.99	8.19	1.20
Drugs	Mefruside	Select	4	13	5	0	0	1	1	0	0	0	1	3.459	0.493	0.507	1.528	1.67	9.61	7.70	-1.91
Drugs	Norfenfluramine	Select	0	10	0	0	0	0	1	0	0	0	1	2.807	0.547	3.246	2.577	0.18	8.60	8.20	-0.40
Drugs	Oxymorphone	Select	1	17	4	0	0	1	4	0	1	3	1	3.000	0.304	1.143	0.778	0.34	2.94	3.50	0.56
Drugs	Phendimetrazine	Select	0	12	1	0	0	0	2	0	0	0	1	2.807	-0.835	1.845	1.998	1.28	6.32	7.84	1.51
Drugs	Phenmetrazine	Select	0	11	1	0	0	0	2	0	0	0	1	2.807	-0.277	1.564	1.462	0.09	5.19	5.95	0.75
Drugs	Trichlormethiazide	Select	4	8	4	0	0	0	2	0	0	0	1	3.459	2.272	0.365	1.364	0.55	4.92	4.21	-0.71
Explosives	2,4-DA-6-NT	Test	2	7	2	0	0	0	1	0	0	0	1	3.170	2.617	1.181	0.717	0.92	3.66	5.88	2.22
Explosives	3-NT	Test	2	7	2	0	0	0	1	0	0	0	1	3.170	-0.672	2.241	2.211	2.45	10.87	9.25	-1.62
Explosives	DMNB	Test	4	6	4	0	0	0	0	0	0	0	0	2.322	-0.484	1.699	1.566	1.82	7.30	9.24	1.94
Explosives	MHN	Test	12	6	18	0	0	0	0	0	0	0	0	3.700	-0.198	2.301	0.031	5.16	14.55	15.98	1.43
Explosives	TMETN	Test	6	5	9	0	0	0	0	0	0	0	0	2.807	-0.288	1.438	0.412	3.08	11.78	11.15	-0.63
Explosives	TNT	Test	6	7	6	0	0	0	1	0	0	0	1	3.700	-0.428	2.359	1.999	1.79	12.25	9.96	-2.29
Drugs	6-MAM	Test	2	19	4	0	0	1	4	0	1	3	1	3.170	-0.342	2.339	1.766	0.50	4.45	6.28	1.83
Drugs	AICAR	Test	1	9	5	0	0	2	0	0	0	0	0	2.807	5.194	-1.404	-2.913	-2.49	1.42	2.76	1.34
Drugs	Andarine	Test	4	19	6	0	0	0	2	0	0	0	2	4.087	1.004	2.447	2.315	3.26	12.40	12.21	-0.19
Drugs	Bumetanide	Test	3	17	5	0	0	0	2	0	0	0	2	4.000	1.815	1.763	2.960	0.21	6.28	7.79	1.51
Drugs	Butizide	Test	4	11	4	0	0	0	2	0	0	0	1	3.459	2.109	0.646	1.577	0.78	6.73	5.24	-1.48
Drugs	Carteolol	Test	1	16	3	0	0	0	2	0	0	0	1	3.000	1.062	1.306	1.283	-0.58	4.17	5.51	1.34

Table A. 3 (Continued): Molecular descriptors and predicted retention time for best ANN model using the Barron & McEneff descriptors.

			Molecular descriptors															Average	Best ANN	
Dataset	Analyte	ANN subset	nDB	nC	nO	nR03	nR04	nR05	nR06	nR07	nR08	nR09	nBnz	Ui	Hy	MLOGP	ALOGP	LogD (pH=7)	t_R^M (min)	t_R^P (min)	Δt_R (min)
Drugs	Chlorthalidone	Test	3	14	4	0	0	1	2	0	0	1	2	4.000	1.959	1.638	1.507	0.41	4.91	6.51	1.60
Drugs	Etacr. acid	Test	3	13	4	0	0	0	1	0	0	0	1	3.322	-0.198	3.010	4.152	-0.24	7.31	8.82	1.52
Drugs	Fenbutrazate	Test	1	23	3	0	0	0	3	0	0	0	2	3.807	-0.843	3.313	4.455	4.45	17.61	14.92	-2.69
Drugs	Fenetylline	Test	2	18	2	0	0	1	2	0	0	1	1	3.807	-0.277	2.444	1.785	-0.07	9.46	8.30	-1.15
Drugs	Fenfluramine	Test	0	12	0	0	0	0	1	0	0	0	1	2.807	-0.229	3.790	3.358	0.48	9.29	9.76	0.46
Drugs	Fluticasone prop.	Test	5	25	5	0	0	1	3	0	0	1	0	2.585	-0.334	3.412	3.947	3.71	16.21	14.45	-1.76
Drugs	Formoterol	Test	1	19	4	0	0	0	2	0	0	0	2	3.807	1.736	1.659	1.929	-0.41	7.71	7.47	-0.24
Drugs	FPCAM	Test	5	24	6	0	0	1	3	0	0	1	0	2.585	0.208	3.146	2.999	-1.11	10.79	9.88	-0.91
Drugs	Mefenorex	Test	0	12	0	0	0	0	1	0	0	0	1	2.807	-0.305	3.390	2.793	0.98	8.50	9.72	1.22
Drugs	Tramadol	Test	0	16	2	0	0	0	2	0	0	0	1	2.807	-0.353	2.318	2.701	0.17	7.12	8.12	0.99

Table A. 4: Molecular descriptors and predicted retention times for ANN and MLR models using Aalizadeh et al. descriptors

			Molecular descriptors								Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	LogD (pH=7)	BLTA96	EEig14r	CIC1	BEHm4	HATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosive	1,2-DNB	Training	1.619	1.60	-3.12	0.000	1.750	2.515	0.038	7.96	7.30	-0.66	6.85	-1.11	7.15	-0.81
Explosive	1,2-DNG	Training	-0.418	0.86	-1.07	0.000	1.423	2.602	0.447	4.42	4.46	0.04	5.47	1.06	6.22	1.81
Explosive	1,3-DNB	Training	1.619	1.55	-3.12	0.000	1.750	2.482	0.331	9.04	8.42	-0.62	7.00	-2.04	7.80	-1.24
Explosive	1,3-DNG	Training	-0.418	0.97	-1.07	0.000	1.423	2.466	0.243	4.02	4.15	0.14	5.17	1.16	5.54	1.53
Explosive	1-MNG	Training	-0.913	-0.43	-0.40	0.000	1.226	2.183	0.200	1.85	1.87	0.02	3.39	1.54	4.00	2.15
Explosive	2,4-DA-6-NT	Training	0.717	0.92	-2.41	0.000	1.215	2.657	0.068	3.66	4.98	1.32	5.78	2.13	6.46	2.81
Explosive	2,6-DA-4-NT	Training	0.717	0.92	-2.41	0.000	1.215	2.694	0.071	3.16	5.00	1.84	5.84	2.68	6.53	3.37
Explosive	2,6-DNT	Training	2.105	2.10	-3.47	0.000	1.382	2.843	0.232	11.04	9.36	-1.68	7.80	-3.24	8.04	-3.00
Explosive	2-Am-4,6-DNT	Training	1.358	1.88	-2.98	-1.000	1.215	2.695	0.508	10.67	8.72	-1.95	7.84	-2.84	7.62	-3.06
Explosive	3,5-DNA	Training	0.872	1.63	-2.62	0.000	1.459	2.642	0.327	9.26	7.29	-1.98	6.68	-2.58	7.58	-1.68
Explosive	4-Am-2,6-DNT	Training	1.358	1.88	-2.98	-1.000	1.215	2.907	0.261	10.33	7.93	-2.40	7.97	-2.36	7.33	-3.01
Explosive	4-NT	Training	2.211	2.45	-3.47	0.000	1.339	2.617	0.062	10.47	9.48	-0.99	7.55	-2.92	7.23	-3.24
Explosive	DADP	Training	1.911	1.23	-2.34	0.000	2.774	2.331	0.032	8.43	7.67	-0.76	7.17	-1.26	6.22	-2.21
Explosive	DEDPU	Training	3.632	3.26	-4.92	0.000	2.691	3.140	0.094	14.81	13.62	-1.19	10.71	-4.10	10.49	-4.31
Explosive	DMDPU	Training	2.935	2.55	-4.43	-0.281	2.587	2.942	0.067	12.44	11.53	-0.91	9.69	-2.75	9.57	-2.86
Explosive	EGDN	Training	0.093	1.51	-1.29	0.000	1.571	2.061	0.303	5.44	5.50	0.06	5.30	-0.15	5.25	-0.19
Explosive	ETN	Training	0.062	3.21	-2.36	0.000	2.289	3.515	0.685	11.60	11.32	-0.28	9.54	-2.06	10.81	-0.79
Explosive	HMDD	Training	1.738	0.00	-1.59	0.000	2.626	2.898	0.010	4.43	5.75	1.32	7.01	2.58	5.67	1.24
Explosive	HMTA	Training	0.086	0.98	-2.34	0.000	3.024	2.878	0.003	1.52	5.53	4.01	7.19	5.67	8.49	6.97
Explosive	HMTD	Training	2.558	0.13	-1.71	-1.122	2.925	3.053	0.025	2.80	6.42	3.63	8.75	5.95	5.21	2.41
Explosive	HND	Training	2.747	4.44	-4.37	0.707	2.737	3.524	0.918	15.57	15.39	-0.18	11.77	-3.80	13.40	-2.17
Explosive	MHN	Training	0.031	5.16	-3.53	0.410	2.843	3.532	0.803	14.55	14.55	0.00	11.09	-3.46	13.89	-0.66
Explosive	NM	Training	0.291	0.10	-0.80	0.000	0.965	0.083	0.000	1.85	0.93	-0.92	0.70	-1.15	0.03	-1.82
Explosive	PA	Training	1.246	-1.40	-2.77	-0.882	1.527	3.019	0.937	4.96	4.25	-0.70	6.71	1.76	8.77	3.81
Explosive	PETN	Training	-0.184	3.64	-2.76	0.000	2.483	3.549	0.610	12.39	12.04	-0.35	9.87	-2.52	11.70	-0.69

Table A. 4 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using Aalizadeh et al. descriptors

			Molecular descriptors								Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	LogD (pH=7)	BLTA96	EEig14r	CIC1	BEHm4	HATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosive	PGDN	Training	0.471	1.87	-1.81	0.000	1.618	2.408	0.315	7.50	6.92	-0.59	6.32	-1.18	6.40	-1.10
Explosive	PYX	Training	4.013	7.04	-5.50	2.209	2.950	3.553	0.583	16.90	16.60	-0.30	12.97	-3.93	14.73	-2.17
Explosive	RDX	Training	0.623	-1.20	-2.63	-0.878	2.156	2.943	1.177	5.95	6.08	0.13	7.08	1.13	10.03	4.08
Explosive	R-Salt	Training	-0.325	-1.18	-0.60	0.000	1.918	2.552	0.598	3.54	2.80	-0.74	4.62	1.08	5.85	2.31
Explosive	TATB	Training	-0.727	2.93	-1.33	-0.517	2.335	3.076	1.972	9.40	10.67	1.27	9.64	0.24	12.01	2.60
Explosive	TATP	Training	2.867	2.16	-3.00	-0.931	3.359	2.335	0.079	11.99	10.31	-1.68	9.40	-2.59	6.79	-5.20
Explosive	Tetryl	Training	1.762	1.25	-3.88	0.000	1.820	3.488	1.203	11.83	10.47	-1.36	9.25	-2.58	12.48	0.66
Explosive	TMETN	Training	0.412	3.08	-2.67	-0.721	2.242	3.313	0.436	11.78	9.99	-1.80	9.60	-2.18	9.71	-2.07
Explosive	TNT	Training	1.999	1.79	-3.59	-0.882	1.608	3.045	0.694	12.25	10.04	-2.21	9.04	-3.21	9.42	-2.82
Drug	6-MAM	Training	1.766	0.50	-3.57	0.490	1.822	3.249	0.126	4.45	7.44	2.99	7.19	2.74	9.19	4.74
Drug	Acebutolol	Training	1.615	-0.73	-2.82	0.133	2.238	3.255	0.039	6.38	5.37	-1.01	6.79	0.41	7.68	1.30
Drug	Acetazolamide	Training	-1.206	-0.58	0.35	0.000	0.711	2.481	0.155	2.15	2.56	0.41	3.13	0.98	9.16	7.01
Drug	AICAR	Training	-2.913	-2.49	0.17	-0.519	1.243	2.963	0.119	1.42	1.59	0.17	2.64	1.22	9.46	8.04
Drug	Alprenolol	Training	2.640	0.34	-3.60	-0.298	2.023	3.240	0.060	11.41	7.37	-4.05	8.14	-3.28	8.96	-2.45
Drug	Altizide	Training	1.486	1.08	-1.80	0.000	1.256	3.753	0.478	6.87	6.04	-0.83	8.24	1.37	8.36	1.49
Drug	Amiloride	Training	-0.077	-0.92	-1.39	-0.970	1.055	2.941	0.231	3.63	2.59	-1.04	5.30	1.67	7.82	4.19
Drug	Aminoglutethimide	Training	1.285	1.27	-2.61	-0.457	1.675	2.879	0.117	4.99	6.34	1.35	7.33	2.34	8.56	3.57
Drug	Anastrazole	Training	2.966	2.68	-4.03	0.000	2.157	3.312	0.137	8.02	11.97	3.95	9.85	1.83	7.62	-0.40
Drug	Andarine	Training	2.315	3.26	-3.68	0.891	1.690	3.377	0.070	12.40	12.39	-0.01	8.95	-3.44	7.93	-4.47
Drug	Beclomethasone	Training	1.905	2.25	-3.19	0.965	2.396	3.281	0.149	11.52	11.86	0.34	8.49	-3.03	9.54	-1.98
Drug	Bendroflumethiazide	Training	2.179	1.48	-2.81	0.240	1.727	3.645	0.442	8.50	8.97	0.46	8.90	0.39	8.24	-0.27
Drug	Benzoylecgonine	Training	1.868	-0.21	-3.40	0.000	2.002	3.187	0.086	4.05	6.07	2.03	7.13	3.09	8.47	4.42
Drug	Betaxolol	Training	2.577	0.43	-3.22	0.000	2.525	3.288	0.051	11.60	8.42	-3.18	8.37	-3.23	7.85	-3.75
Drug	Bisoprolol	Training	2.031	-0.22	-2.83	0.000	2.757	3.274	0.046	9.59	7.22	-2.37	7.83	-1.76	8.35	-1.25
Drug	Budesonide	Training	2.132	3.02	-3.24	1.115	2.750	3.337	0.084	15.29	13.81	-1.48	9.28	-6.01	9.84	-5.45
Drug	Bumetanide (-)	Training	2.960	0.21	-2.99	0.230	1.746	3.390	0.248	6.25	7.42	1.17	7.95	1.70	6.61	0.36
Drug	Bunolol	Training	2.358	0.15	-3.55	0.000	2.233	3.264	0.035	7.59	7.38	-0.21	7.85	0.26	7.61	0.02

Table A. 4 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using Aalizadeh et al. descriptors

			Molecular descriptors								Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	LogD (pH=7)	BLTA96	EEig14r	CIC1	BEHm4	HATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drug	Carphedon	Training	0.254	0.31	-2.07	-0.683	1.629	2.808	0.114	5.16	3.67	-1.50	6.22	1.06	6.53	1.36
Drug	Carvedilol	Training	4.015	2.67	-3.43	1.369	2.476	3.479	0.050	13.30	14.89	1.59	9.76	-3.54	5.95	-7.35
Drug	Chlorexolone	Training	2.348	1.49	-3.20	0.000	1.782	3.450	0.126	9.74	8.58	-1.17	8.67	-1.07	6.25	-3.49
Drug	Chlorothiazide	Training	0.450	-0.10	-1.42	-0.543	0.935	3.202	1.129	2.08	3.19	1.12	6.79	4.71	8.29	6.22
Drug	Chlorthalidone	Training	1.507	0.41	-2.87	0.109	1.456	3.363	0.289	4.91	5.99	1.08	7.25	2.34	10.64	5.73
Drug	Clobenzorex	Training	4.314	2.89	-5.64	-0.266	2.307	3.234	0.090	12.70	13.68	0.98	10.89	-1.81	10.74	-1.96
Drug	Clopamide	Training	2.318	1.75	-3.24	0.000	1.865	3.406	0.093	6.01	9.02	3.01	8.79	2.79	10.30	4.29
Drug	Cyclopenthiazide	Training	2.118	1.51	-2.41	0.000	1.918	3.663	0.430	9.61	8.77	-0.84	9.18	-0.43	5.41	-4.20
Drug	Danazol	Training	4.875	4.33	-5.38	0.738	2.431	3.201	0.079	16.65	15.50	-1.15	11.36	-5.29	9.02	-7.64
Drug	Desonide	Training	0.962	2.79	-3.04	1.106	2.623	3.306	0.095	12.50	12.14	-0.36	8.44	-4.06	10.24	-2.26
Drug	Diacetolol	Training	0.492	-1.40	-2.34	0.084	2.012	3.253	0.052	3.62	3.29	-0.32	5.68	2.07	2.94	-0.68
Drug	Dichlorphenamide	Training	0.815	0.89	-1.43	-0.940	1.364	3.312	0.956	4.19	5.50	1.31	8.21	4.02	7.45	3.26
Drug	Esmolol	Training	1.978	-0.42	-3.07	-0.004	2.233	3.251	0.069	8.36	6.14	-2.22	7.29	-1.07	6.76	-1.60
Drug	Etacr. acid	Training	4.152	-0.24	-4.24	-0.470	1.150	3.305	0.559	7.31	7.90	0.59	8.49	1.18	7.83	0.52
Drug	Etamivan	Training	1.658	1.44	-2.94	-0.788	1.928	2.899	0.097	5.75	7.10	1.35	8.08	2.33	7.27	1.51
Drug	Famprofazone	Training	4.847	1.18	-5.62	0.883	2.804	3.426	0.063	17.86	15.37	-2.49	9.92	-7.94	11.29	-6.57
Drug	Fencamfamin	Training	3.086	0.39	-4.75	-0.570	2.394	2.854	0.078	10.22	9.12	-1.10	8.36	-1.86	9.27	-0.95
Drug	Fenetylline	Training	1.785	-0.07	-3.67	0.643	2.065	3.356	0.053	9.46	7.28	-2.18	7.04	-2.42	9.51	0.06
Drug	Fludrocortisone	Training	1.346	1.54	-2.96	0.780	2.593	3.193	0.099	9.59	10.00	0.41	7.87	-1.72	9.32	-0.27
Drug	Flumethasone	Training	1.480	1.74	-3.19	1.056	2.327	3.243	0.125	11.13	10.69	-0.44	7.78	-3.36	9.66	-1.47
Drug	FPCAM	Training	2.999	-1.11	-4.38	1.256	2.437	3.337	0.131	10.79	10.27	-0.52	6.90	-3.89	10.36	-0.43
Drug	Fulvestrant	Training	8.412	7.56	-7.30	2.108	3.542	3.598	0.074	18.08	17.49	-0.59	15.68	-2.40	8.68	-9.40
Drug	Furosemide	Training	1.714	-0.74	-1.66	0.000	0.886	3.434	0.378	3.36	3.16	-0.21	6.35	2.98	9.68	6.32
Drug	Gestrinone	Training	4.132	3.26	-5.10	0.549	2.234	3.172	0.072	14.00	14.08	0.08	10.26	-3.74	5.04	-8.96
Drug	Hydroflumeth.	Training	0.316	0.23	-1.15	-0.300	1.098	3.186	0.696	2.63	3.70	1.07	6.50	3.87	6.99	4.36
Drug	Labetalol	Training	2.330	0.53	-3.90	0.189	1.975	3.273	0.069	9.84	7.95	-1.89	7.82	-2.02	8.30	-1.54
Drug	Mefruside	Training	1.528	1.67	-1.74	0.000	1.876	3.668	0.400	9.61	7.96	-1.65	8.90	-0.71	7.74	-1.87

Table A. 4 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using Aalizadeh et al. descriptors

			Molecular descriptors								Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	LogD (pH=7)	BLTA96	EEig14r	CIC1	BEHm4	HATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drug	Methylprednisolone	Training	1.446	1.97	-2.98	0.901	2.558	3.267	0.086	11.44	10.88	-0.56	8.18	-3.26	9.44	-2.00
Drug	Metoprolol	Training	1.757	-0.59	-2.88	-0.122	2.289	3.205	0.057	7.27	5.52	-1.75	7.11	-0.16	8.85	1.58
Drug	Nadoxolol	Training	1.369	1.63	-3.20	-0.295	1.626	3.163	0.070	9.26	7.28	-1.98	7.90	-1.36	8.42	-0.84
Drug	Nikethamide	Training	0.791	0.54	-2.25	0.000	1.823	2.899	0.082	3.61	4.99	1.38	6.41	2.80	7.00	3.38
Drug	Norfenfluramine	Training	2.577	0.18	-4.48	-1.000	1.549	2.685	0.118	8.60	7.01	-1.59	7.40	-1.20	8.14	-0.46
Drug	Ostarine	Training	2.315	3.26	-3.68	0.891	1.690	3.377	0.070	13.34	12.39	-0.95	8.95	-4.39	7.75	-5.59
Drug	Oxprenolol	Training	2.232	-0.15	-3.06	-0.247	2.070	3.248	0.059	9.46	6.11	-3.35	7.61	-1.85	7.84	-1.62
Drug	Oxymorphone	Training	0.778	0.34	-2.37	0.094	1.771	3.126	0.181	2.94	5.23	2.29	6.62	3.68	7.75	4.81
Drug	Pemoline	Training	1.336	0.50	-2.41	-1.033	1.201	2.617	0.148	3.15	5.10	1.95	6.52	3.38	5.55	2.40
Drug	Pentazocine	Training	4.348	1.89	-5.01	0.000	2.435	3.098	0.076	8.59	13.02	4.43	9.92	1.32	9.60	1.01
Drug	Pentetrazol	Training	0.918	0.22	-2.90	0.000	1.899	2.376	0.008	2.86	4.23	1.37	5.52	2.66	6.99	4.13
Drug	Phendimetrazine	Training	1.998	1.28	-3.07	-1.000	2.137	2.891	0.073	6.32	7.28	0.96	8.43	2.10	7.19	0.87
Drug	Phenmetrazine	Training	1.462	0.09	-2.79	-1.107	1.848	2.779	0.070	5.19	4.69	-0.50	7.09	1.90	6.53	1.34
Drug	Pindolol	Training	1.926	-0.65	-2.54	-0.260	1.753	3.193	0.050	4.86	4.51	-0.35	6.78	1.92	7.69	2.83
Drug	Piretanide	Training	2.446	0.07	-2.60	0.335	1.705	3.514	0.349	5.12	6.72	1.60	7.76	2.64	7.87	2.75
Drug	Prenylamine	Training	5.711	2.60	-6.71	0.561	3.077	3.429	0.075	15.88	16.68	0.80	11.75	-4.13	12.62	-3.26
Drug	Propylhexedrine	Training	2.739	-0.30	-3.73	0.000	3.186	2.747	0.049	8.76	9.15	0.39	7.70	-1.07	8.56	-0.20
Drug	Salmeterol	Training	4.226	1.62	-4.10	1.290	2.912	3.408	0.053	15.58	15.40	-0.18	9.54	-6.05	8.23	-7.35
Drug	Sotalol	Training	0.967	-1.97	-1.94	-0.208	1.845	3.130	0.060	2.91	2.86	-0.05	5.41	2.50	7.59	4.68
Drug	Tapentadol	Training	3.441	0.28	-4.27	-0.790	2.437	2.902	0.096	7.33	9.26	1.92	8.69	1.36	8.28	0.94
Drug	Testolactone	Training	3.123	3.07	-4.71	0.276	2.560	3.172	0.085	9.17	13.11	3.94	10.08	0.91	8.71	-0.46
Drug	Timolol	Training	1.128	-0.71	-2.45	-0.006	2.542	3.328	0.158	7.95	5.64	-2.31	7.08	-0.86	7.61	-0.33
Drug	Tramadol	Training	2.701	0.17	-3.55	-0.100	2.555	3.035	0.086	7.12	8.49	1.37	8.02	0.90	8.31	1.19
Drug	Triamcinolone	Training	0.356	0.92	-2.10	0.956	2.217	3.204	0.124	7.24	7.42	0.18	6.57	-0.66	8.50	1.26
Drug	Trichlormethiazide	Training	1.364	0.55	-1.59	-0.149	1.000	3.753	0.721	4.92	4.56	-0.36	7.94	3.02	8.32	3.40
Drug	Xipamide	Training	2.812	1.15	-3.61	0.000	1.302	3.424	0.195	5.01	8.12	3.11	8.36	3.35	10.58	5.57
Explosive	2,4-DNT	Select	2.105	2.10	-3.47	0.000	1.382	2.693	0.518	11.21	10.10	-1.11	7.80	-3.41	8.52	-2.69

Table A. 4 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using Aalizadeh et al. descriptors

			Molecular descriptors								Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	LogD (pH=7)	BLTA96	EEig14r	CIC1	BEHm4	HATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosive	3,4-DNT	Select	2.105	2.15	-3.47	0.000	1.382	2.673	0.148	10.13	9.18	-0.96	7.50	-2.63	7.60	-2.54
Explosive	DMNB	Select	1.566	1.82	-2.93	0.000	2.626	2.714	0.000	7.30	8.22	0.92	7.89	0.59	7.84	0.53
Explosive	HMX	Select	0.831	-1.71	-3.59	-0.055	2.571	3.468	0.863	4.64	7.82	3.18	7.25	2.62	12.08	7.44
Explosive	NB	Select	1.724	1.81	-3.12	0.000	1.801	2.397	0.010	8.00	7.56	-0.43	6.86	-1.13	6.90	-1.10
Explosive	TNB	Select	1.513	1.36	-3.24	-0.882	1.918	2.984	0.844	9.81	9.48	-0.33	8.77	-1.04	9.71	-0.10
Drug	Atenolol	Select	0.669	-2.19	-2.15	-0.085	1.855	3.225	0.057	2.73	2.70	-0.03	5.21	2.48	7.66	4.93
Drug	Benzthiazide	Select	2.498	0.99	-3.02	0.298	1.651	3.755	0.330	9.39	8.22	-1.18	8.74	-0.65	13.59	4.20
Drug	Bumetanide (+)	Select	2.960	0.21	-2.99	0.230	1.746	3.390	0.248	6.28	7.42	1.14	7.95	1.67	7.63	1.34
Drug	Bupropion	Select	3.227	2.69	-4.45	-0.769	2.121	3.115	0.133	9.54	11.85	2.31	10.22	0.68	9.19	-0.35
Drug	Butizide	Select	1.577	0.78	-1.88	0.000	1.601	3.485	0.525	6.73	6.37	-0.36	7.99	1.26	9.70	2.97
Drug	Carteolol	Select	1.283	-0.58	-2.54	0.000	2.074	3.284	0.046	4.17	4.71	0.54	6.73	2.56	8.45	4.28
Drug	Desacetyl deflazacort	Select	1.224	1.84	-3.15	1.104	2.350	3.254	0.097	10.85	10.60	-0.25	7.69	-3.16	9.80	-1.05
Drug	Efaproxiral	Select	3.858	0.38	-4.36	0.418	2.153	3.321	0.057	7.28	10.31	3.03	8.52	1.23	9.46	2.18
Drug	Epitizide	Select	1.876	1.01	-2.01	0.000	1.210	3.753	0.586	6.68	6.23	-0.45	8.45	1.77	7.66	0.99
Drug	Fenbutrazate	Select	4.455	4.45	-4.54	0.905	2.771	3.424	0.073	17.61	16.48	-1.13	11.65	-5.96	10.33	-7.28
Drug	Fenfluramine	Select	3.358	0.48	-5.02	-0.788	1.963	2.984	0.086	9.29	9.11	-0.19	8.60	-0.69	9.24	-0.05
Drug	Hydrochloroth.	Select	0.038	0.00	-0.68	-0.543	1.040	3.212	0.894	2.13	2.86	0.73	6.50	4.38	6.09	3.97
Drug	Mefenorex	Select	2.793	0.98	-4.62	-1.000	2.070	2.944	0.118	8.50	8.94	0.44	8.81	0.31	9.11	0.61
Drug	Polythiazide	Select	2.412	1.59	-2.28	0.000	1.356	3.754	0.434	8.84	7.86	-0.98	9.08	0.24	11.80	2.97
Drug	Probenecid	Select	2.821	0.13	-3.21	-0.011	2.034	3.155	0.170	5.30	7.65	2.35	7.82	2.52	8.08	2.78
Drug	Strychnine	Select	1.146	0.21	-4.13	0.938	1.981	3.374	0.149	8.14	8.05	-0.09	6.79	-1.35	11.09	2.94
Explosive	2-NT	Test	2.211	2.45	-3.47	0.000	1.339	2.467	0.017	10.29	9.27	-1.02	7.28	-3.01	6.90	-3.39
Explosive	3-NT	Test	2.211	2.45	-3.47	0.000	1.339	2.427	0.077	10.87	9.48	-1.39	7.27	-3.60	6.99	-3.88
Explosive	DPA	Test	3.380	3.12	-4.62	-1.000	2.852	2.617	0.092	13.97	12.03	-1.94	10.48	-3.48	9.05	-4.92
Explosive	NG	Test	0.078	2.41	-1.81	-0.982	1.931	3.111	0.581	8.97	8.09	-0.87	8.70	-0.26	8.20	-0.77
Explosive	NQ	Test	-0.532	-2.56	-0.31	0.000	0.614	1.711	0.000	1.49	1.01	-0.49	0.87	-0.62	1.62	0.13
Explosive	SHN	Test	0.031	5.16	-3.53	0.410	2.843	3.532	0.779	14.55	14.54	-0.01	11.08	-3.48	13.83	-0.72

Table A. 4 (Continued): Molecular descriptors and predicted retention times for ANN and MLR models using Aalizadeh et al. descriptors

			Molecular descriptors								Best ANN		Original MLR		Optimised MLR	
Dataset	Analyte	ANN subset	ALOGP	LogD (pH=7)	BLTA96	EEig14r	CIC1	BEHm4	HATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosive	TEGDN	Test	-0.169	1.46	-1.58	-0.934	2.590	3.132	0.111	7.61	4.92	-2.69	8.10	0.49	7.36	-0.25
Drug	Bambuterol	Test	1.727	-0.93	-2.26	0.000	2.587	3.434	0.048	8.41	5.63	-2.79	7.32	-1.10	8.16	-0.26
Drug	Beta\dexameth.	Test	1.642	1.92	-3.08	0.972	2.396	3.213	0.108	10.89	10.91	0.02	8.01	-2.89	9.32	-1.57
Drug	Canrenone	Test	3.379	2.50	-5.38	0.762	2.762	3.224	0.067	14.71	13.91	-0.80	9.78	-4.92	8.70	-6.01
Drug	Celiprolol	Test	1.597	-0.41	-2.93	0.377	2.638	3.318	0.031	7.66	6.89	-0.77	7.22	-0.44	6.55	-1.12
Drug	Clobetasol	Test	2.845	2.74	-4.09	0.866	2.340	3.380	0.129	12.94	13.35	0.41	9.47	-3.47	10.30	-2.64
Drug	Deflazacort	Test	1.603	2.43	-3.53	1.291	2.498	3.413	0.083	13.62	12.70	-0.92	8.49	-5.13	10.60	-3.02
Drug	Fluticasone prop.	Test	3.947	3.71	-4.64	1.363	2.495	3.441	0.144	16.21	16.09	-0.12	10.51	-5.69	10.97	-5.24
Drug	Formoterol	Test	1.929	-0.41	-2.89	0.464	1.932	3.242	0.074	7.71	5.97	-1.73	6.69	-1.02	6.98	-0.73
Drug	Isometheptene	Test	2.405	-0.03	-3.68	0.000	2.465	2.609	0.046	6.99	7.41	0.42	6.94	-0.05	7.96	0.97
Drug	Oxycodone	Test	1.029	0.15	-2.61	0.456	2.024	3.128	0.131	4.60	6.02	1.41	6.54	1.94	8.18	3.58
Drug	Pethidine	Test	2.446	1.56	-3.75	-0.245	2.339	3.023	0.103	8.41	9.54	1.13	8.72	0.31	8.77	0.36
Drug	Prolintane	Test	4.043	1.36	-4.75	-0.691	2.776	2.970	0.079	9.25	12.07	2.82	9.97	0.72	9.11	-0.14
Drug	Propranolol	Test	2.540	0.81	-3.76	-0.087	2.158	3.225	0.045	11.61	8.36	-3.24	8.32	-3.29	7.59	-4.02
Drug	Terbutaline	Test	1.254	-1.94	-2.36	-0.724	1.887	2.856	0.071	2.83	2.85	0.02	5.57	2.74	8.11	5.28
Drug	Triamterene	Test	1.017	1.10	-3.93	-0.043	1.783	2.970	0.135	6.56	6.93	0.37	7.17	0.61	9.44	2.89

Table A. 5: Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Average	Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	1,2-DNB	Training	1.6	3.557	0.000	0.144	0.000	12	0.000	0	0.000	40	0.000	7.96	8.54	0.58	8.16	0.20
Explosives	1,3-DNB	Training	1.55	3.557	0.000	0.096	0.000	8	0.000	0	0.000	30	0.000	9.04	9.42	0.38	8.87	-0.17
Explosives	1,3-DNG	Training	0.97	0.027	0.000	0.130	0.602	8	0.000	1	0.000	14	0.000	4.02	5.57	1.55	6.60	2.58
Explosives	2,4-DA-6-NT	Training	0.92	1.395	0.000	0.086	0.000	8	0.000	4	0.000	54	0.000	3.66	6.78	3.12	6.53	2.87
Explosives	2-Am-4,6-DNT	Training	1.88	3.055	0.000	0.086	0.000	40	0.000	2	0.000	92	0.837	10.67	8.96	-1.71	8.63	-2.04
Explosives	3,4-DNT	Training	2.15	5.014	0.000	0.113	0.000	24	0.000	0	0.000	66	2.328	10.13	9.21	-0.93	9.38	-0.75
Explosives	3-NT	Training	2.45	5.02	0.000	0.129	0.000	0	0.000	0	0.000	13	0.000	10.87	10.99	0.12	9.86	-1.01
Explosives	4-Am-2,6-DNT	Training	1.88	3.055	0.000	0.083	0.000	40	0.000	2	0.000	108	0.000	10.33	9.61	-0.72	8.82	-1.51
Explosives	4-NT	Training	2.45	5.02	0.000	0.160	0.000	0	0.000	0	0.000	10	2.768	10.47	8.74	-1.74	9.00	-1.47
Explosives	DADP	Training	1.23	1.225	0.000	0.209	0.000	0	0.000	0	0.000	4	0.000	8.43	5.46	-2.98	5.98	-2.45
Explosives	DEDPU	Training	3.26	13.651	2.157	0.103	2.398	636	0.780	0	0.000	270	0.794	14.81	14.58	-0.23	13.79	-1.02
Explosives	DMDPU	Training	2.55	10.254	2.495	0.096	2.197	472	0.682	0	0.000	188	1.039	12.44	12.41	-0.03	12.31	-0.13
Explosives	DMNB	Training	1.82	2.887	0.000	0.210	0.000	0	0.000	0	0.000	16	0.000	7.30	7.29	-0.01	7.17	-0.13
Explosives	DPA	Training	3.12	11.526	1.095	0.210	0.693	136	0.000	1	0.000	44	0.000	13.97	13.00	-0.97	11.22	-2.75
Explosives	EGDN	Training	1.51	0.003	0.000	0.137	0.000	0	0.000	0	0.000	6	0.700	5.44	5.97	0.53	7.02	1.58
Explosives	HMDD	Training	0	0.129	0.000	0.250	0.000	60	0.000	0	0.000	28	0.000	4.43	2.45	-1.98	3.57	-0.86
Explosives	HMTA	Training	0.98	1.225	0.000	0.335	0.000	0	0.000	0	0.000	72	0.000	1.52	3.41	1.89	3.76	2.24
Explosives	HMTD	Training	0.13	0.226	0.000	0.223	0.000	138	0.000	0	0.000	42	0.000	2.80	3.07	0.27	4.16	1.36
Explosives	HMX	Training	-1.71	5.556	0.000	0.177	0.976	384	0.443	0	0.000	248	0.260	4.64	3.26	-1.38	4.22	-0.42
Explosives	HND	Training	4.44	9.868	5.059	0.114	2.812	2992	0.623	1	0.004	1300	0.739	15.57	15.24	-0.33	14.47	-1.10
Explosives	MHN	Training	5.16	5.296	7.066	0.044	2.528	788	0.835	0	0.006	702	0.983	14.55	15.56	1.01	14.73	0.18
Explosives	NB	Training	1.81	3.562	0.000	0.129	0.000	0	0.000	0	0.000	4	0.000	8.00	9.18	1.18	8.66	0.66
Explosives	NG	Training	2.41	0.331	0.000	0.089	0.602	20	0.000	0	0.000	50	0.447	8.97	8.51	-0.47	9.35	0.38
Explosives	NM	Training	0.1	0.187	0.000	0.334	0.000	0	0.000	0	0.000	0	0.000	1.85	1.74	-0.11	2.50	0.65

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	NQ	Training	-2.56	0.837	0.000	0.148	0.000	0	0.000	4	0.000	0	0.000	1.49	0.32	-1.17	1.33	-0.16
Explosives	PA	Training	-1.4	2.38	0.000	0.101	0.000	96	0.000	1	0.000	160	0.000	4.96	3.68	-1.28	4.52	-0.44
Explosives	PETN	Training	3.64	2.355	0.000	0.072	1.785	144	0.559	0	0.000	168	0.000	12.39	10.85	-1.54	11.63	-0.76
Explosives	PGDN	Training	1.87	0.331	0.000	0.130	0.000	0	0.000	0	0.000	18	0.608	7.50	6.88	-0.62	7.69	0.19
Explosives	RDX	Training	-1.2	1.963	0.000	0.183	0.000	72	0.000	0	0.000	96	0.000	5.95	2.63	-3.32	3.68	-2.27
Explosives	R-Salt	Training	-1.18	0.403	0.000	0.235	0.000	18	0.000	0	0.000	36	0.000	3.54	1.10	-2.44	2.51	-1.03
Explosives	SHN	Training	5.16	5.296	4.133	0.045	2.528	788	0.835	0	0.006	702	0.983	14.55	14.89	0.34	14.38	-0.17
Explosives	TEGDN	Training	1.46	0.12	9.367	0.108	1.383	52	0.080	0	0.008	18	0.982	7.61	9.21	1.60	8.99	1.38
Explosives	Tetryl	Training	1.25	7.045	0.000	0.113	0.602	360	0.220	0	0.000	362	0.269	11.83	9.87	-1.97	9.28	-2.55
Explosives	TMETN	Training	3.08	2.067	0.000	0.092	1.247	60	0.000	0	0.000	84	0.000	11.78	10.59	-1.19	10.97	-0.81
Explosives	TNB	Training	1.36	4.022	0.000	0.069	0.000	72	0.000	0	0.000	96	0.000	9.81	9.96	0.15	9.21	-0.60
Explosives	TNT	Training	1.79	5.563	0.000	0.083	0.000	96	0.000	0	0.000	160	0.331	12.25	10.94	-1.31	9.90	-2.36
Drugs	Acebutolol	Training	-0.73	2.526	15.143	0.080	2.822	590	0.761	3	0.006	347	1.375	6.38	7.20	0.82	7.62	1.24
Drugs	Acetazolamide	Training	-0.58	2.505	2.238	0.278	0.000	6	0.000	3	0.000	69	1.415	2.15	3.16	1.01	2.73	0.58
Drugs	AICAR	Training	-2.49	1.97	3.657	0.120	1.107	449	0.109	7	0.000	346	1.123	1.42	2.00	0.58	2.31	0.89
Drugs	Alprenolol	Training	0.34	5.619	8.333	0.110	2.371	222	0.676	2	0.012	103	1.638	11.41	8.65	-2.76	8.06	-3.36
Drugs	Altizide	Training	1.08	0.323	6.583	0.203	2.517	747	0.501	4	0.010	681	0.973	6.87	6.46	-0.41	6.10	-0.77
Drugs	Amiloride	Training	-0.92	0.025	1.275	0.133	0.577	76	0.000	8	0.000	130	0.821	3.63	2.61	-1.02	2.66	-0.97
Drugs	Anastrazole	Training	2.68	7.813	2.616	0.146	2.578	732	0.754	0	0.000	420	1.490	8.02	10.21	2.19	11.26	3.24
Drugs	Beclomethasone	Training	2.25	3.848	10.683	0.135	2.751	5110	0.967	3	0.009	2574	1.319	11.52	10.30	-1.22	10.56	-0.96
Drugs	Benzoylecgonine	Training	-0.21	4.71	6.634	0.170	2.570	1164	0.608	1	0.012	537	0.622	4.05	7.00	2.95	6.73	2.68
Drugs	Benzthiazide	Training	0.99	3.216	12.931	0.169	3.084	1835	0.648	3	0.008	1001	0.930	9.39	7.94	-1.45	8.59	-0.80
Drugs	Beta\dexameth.	Training	1.92	3.439	9.534	0.119	2.724	5110	0.966	3	0.009	2574	1.035	10.89	10.02	-0.87	10.20	-0.69
Drugs	Betaxolol	Training	0.43	3.964	10.256	0.099	2.772	504	0.799	2	0.008	147	1.334	11.60	8.54	-3.06	8.54	-3.06
Drugs	Bisoprolol	Training	-0.22	2.545	11.248	0.101	2.859	458	0.764	2	0.010	148	1.061	9.59	7.60	-1.99	7.52	-2.07
Drugs	Budesonide	Training	3.02	4.058	25.012	0.103	3.174	7476	1.007	2	0.009	3292	0.989	15.29	14.96	-0.33	14.05	-1.24

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Bumetanide	Training	0.21	3.108	14.148	0.115	2.967	1105	0.761	4	0.009	598	0.932	6.28	7.83	1.55	7.99	1.71
Drugs	Bunolol	Training	0.15	5.397	17.900	0.130	2.638	650	0.761	2	0.014	312	1.654	7.59	8.74	1.15	8.58	0.99
Drugs	Butizide	Training	0.78	0.417	5.844	0.182	2.159	624	0.511	4	0.000	673	0.745	6.73	4.26	-2.47	6.02	-0.71
Drugs	Canrenone	Training	2.5	17.257	9.204	0.141	2.690	4256	0.932	0	0.008	1734	1.010	14.71	14.94	0.23	13.92	-0.79
Drugs	Carteolol	Training	-0.58	1.706	15.700	0.127	2.606	642	0.761	3	0.009	305	1.572	4.17	6.99	2.82	6.75	2.58
Drugs	Carvedilol	Training	2.67	4.819	18.370	0.105	3.018	3136	0.972	3	0.004	970	0.989	13.30	13.82	0.52	12.66	-0.64
Drugs	Celiprolol	Training	-0.41	2.876	19.924	0.085	3.106	935	0.802	3	0.008	485	1.415	7.66	8.33	0.67	8.66	1.00
Drugs	Chlorexolone	Training	1.49	3.875	6.391	0.187	2.472	1121	0.539	2	0.012	625	0.722	9.74	7.95	-1.79	7.96	-1.78
Drugs	Chlorthalidone	Training	0.41	2.681	0.333	0.194	2.347	1439	0.492	4	0.000	858	1.285	4.91	4.46	-0.45	5.61	0.70
Drugs	Clobenzorex	Training	2.89	19.458	2.784	0.185	2.601	442	0.518	1	0.006	138	1.208	12.70	12.65	-0.05	12.99	0.29
Drugs	Clopamide	Training	1.75	4.034	12.067	0.165	2.952	962	0.586	3	0.012	502	0.623	6.01	8.71	2.70	9.14	3.13
Drugs	Deflazacort	Training	2.43	5.312	23.340	0.089	3.264	7576	1.086	1	0.010	3464	0.986	13.62	14.71	1.09	13.88	0.26
Drugs	Desacetyl deflazacort	Training	1.84	3.67	13.903	0.107	2.962	6704	0.993	2	0.010	2995	1.067	10.85	11.82	0.97	11.43	0.58
Drugs	Diacetolol	Training	-1.4	1.229	8.363	0.082	2.557	431	0.730	3	0.006	288	1.525	3.62	5.08	1.46	5.48	1.86
Drugs	Dichlorphenamide	Training	0.89	0.039	0.870	0.292	0.000	78	0.000	4	0.000	191	0.611	4.19	3.09	-1.10	3.06	-1.13
Drugs	Efaproxiral	Training	0.38	9.766	9.435	0.111	2.915	1148	0.761	2	0.011	592	1.291	7.28	9.24	1.96	9.51	2.23
Drugs	Epitizide	Training	1.01	0.613	7.292	0.191	2.507	923	0.331	4	0.009	839	0.981	6.68	6.50	-0.19	6.54	-0.14
Drugs	Esmolol	Training	-0.42	3.372	13.074	0.097	2.525	386	0.761	2	0.012	125	1.522	8.36	7.96	-0.40	7.46	-0.90
Drugs	Etacr. acid	Training	-0.24	9.061	3.865	0.252	1.761	280	0.221	1	0.007	230	1.589	7.31	7.91	0.60	6.41	-0.90
Drugs	Famprofazone	Training	1.18	19.259	24.626	0.105	3.280	2378	0.879	0	0.006	1075	0.994	17.86	16.67	-1.19	14.93	-2.93
Drugs	Fenbutrazate	Training	4.45	10.976	14.590	0.112	3.058	1933	0.877	0	0.006	606	0.955	17.61	16.88	-0.73	15.47	-2.14
Drugs	Fenetylline	Training	-0.07	5.972	16.905	0.128	2.807	1762	0.882	1	0.008	779	0.468	9.46	9.69	0.23	9.36	-0.10
Drugs	Fenfluramine	Training	0.48	14.367	0.881	0.134	1.375	98	0.288	1	0.000	122	0.896	9.29	10.93	1.64	9.90	0.61
Drugs	Flumethasone	Training	1.74	3.848	13.027	0.115	2.754	5828	0.963	3	0.009	2905	1.200	11.13	10.87	-0.26	10.73	-0.40
Drugs	Formoterol	Training	-0.41	2.753	21.550	0.111	2.889	1227	0.799	4	0.008	391	0.929	7.71	8.39	0.68	8.16	0.45
Drugs	FPCAM	Training	-1.11	9.9	15.058	0.098	3.090	7218	0.964	2	0.008	3594	1.082	10.79	10.91	0.12	10.19	-0.60

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors											Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Fulvestrant	Training	7.56	36.885	41.129	0.105	3.539	6581	1.232	2	0.006	2507	0.733	18.08	18.65	0.57	20.47	2.39
Drugs	Furosemide	Training	-0.74	0.189	5.187	0.139	2.418	586	0.302	4	0.010	426	0.843	3.36	5.38	2.02	4.76	1.40
Drugs	Hydrochloroth.	Training	0	0.299	0.000	0.225	0.000	270	0.000	4	0.000	341	0.821	2.13	2.56	0.43	2.99	0.86
Drugs	Hydroflumeth.	Training	0.23	0.006	0.000	0.153	0.000	603	0.061	4	0.000	566	0.710	2.63	3.32	0.69	4.21	1.58
Drugs	Isometheptene	Training	-0.03	5.983	0.603	0.129	0.000	0	0.000	1	0.000	5	2.000	6.99	6.97	-0.02	6.68	-0.31
Drugs	Mefenorex	Training	0.98	11.489	0.044	0.165	1.868	74	0.142	1	0.012	42	1.570	8.50	9.10	0.60	8.80	0.30
Drugs	Mefruside	Training	1.67	0.257	7.805	0.179	2.658	971	0.623	2	0.010	685	0.705	9.61	7.42	-2.19	7.75	-1.86
Drugs	Methylprednisolone	Training	1.97	3.051	14.854	0.117	2.677	4367	0.956	3	0.011	2176	1.003	11.44	10.70	-0.74	10.54	-0.90
Drugs	Metoprolol	Training	-0.59	2.732	6.922	0.103	2.397	272	0.761	2	0.011	95	1.497	7.27	7.07	-0.20	6.30	-0.97
Drugs	Nikethamide	Training	0.54	1.041	0.001	0.156	0.000	48	0.164	0	0.000	38	0.421	3.61	4.52	0.91	5.66	2.05
Drugs	Norfenfluramine	Training	0.18	10.54	0.004	0.165	0.000	24	0.000	2	0.000	83	1.271	8.60	8.87	0.27	7.30	-1.30
Drugs	Oxprenolol	Training	-0.15	3.351	3.473	0.094	2.412	263	0.678	2	0.010	113	1.248	9.46	7.35	-2.11	6.78	-2.68
Drugs	Oxymorphone	Training	0.34	1.308	0.553	0.171	0.375	3325	0.640	2	0.000	1627	1.498	2.94	3.46	0.52	5.03	2.09
Drugs	Pentazocine	Training	1.89	14.263	5.471	0.122	2.347	1168	0.761	1	0.019	655	1.421	8.59	9.41	0.82	10.75	2.16
Drugs	Pentetrazol	Training	0.22	2.776	0.000	0.293	0.000	12	0.000	0	0.000	20	0.000	2.86	3.82	0.96	3.95	1.09
Drugs	Pethidine	Training	1.56	6.362	3.045	0.124	1.386	352	0.640	0	0.000	243	1.249	8.41	8.87	0.46	9.44	1.03
Drugs	Piretanide	Training	0.07	1.88	4.319	0.114	2.592	1527	0.743	3	0.012	748	1.088	5.12	6.45	1.33	6.35	1.23
Drugs	Polythiazide	Training	1.59	1.104	10.946	0.170	2.507	1159	0.494	3	0.009	948	1.020	8.84	7.78	-1.06	8.22	-0.62
Drugs	Prenylamine	Training	2.6	29.979	7.451	0.125	3.135	1368	0.799	1	0.001	394	2.155	15.88	16.77	0.89	16.07	0.19
Drugs	Probenecid	Training	0.13	3.915	5.937	0.134	1.920	288	0.602	1	0.000	226	0.789	5.30	6.08	0.78	7.38	2.08
Drugs	Prolintane	Training	1.36	12.371	5.665	0.142	1.386	231	0.414	0	0.000	105	0.500	9.25	12.52	3.27	11.16	1.91
Drugs	Propranolol	Training	0.81	6.42	6.232	0.123	2.427	458	0.626	2	0.009	192	1.718	11.61	8.99	-2.62	8.63	-2.98
Drugs	Salmeterol	Training	1.62	8.248	21.373	0.104	2.991	1412	1.034	4	0.004	352	1.053	15.58	13.29	-2.29	12.03	-3.55
Drugs	Tapentadol	Training	0.28	9.244	0.586	0.114	1.363	111	0.624	1	0.000	138	1.690	7.33	8.03	0.70	8.18	0.85
Drugs	Testolactone	Training	3.07	12.144	0.622	0.152	2.045	2200	0.744	0	0.021	1015	1.351	9.17	9.11	-0.06	10.82	1.65
Drugs	Timolol	Training	-0.71	1.496	20.367	0.121	2.594	572	0.761	2	0.010	258	1.337	7.95	7.58	-0.37	7.41	-0.54

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Tramadol	Training	0.17	5.374	4.331	0.117	1.711	492	0.717	1	0.000	315	1.110	7.12	6.59	-0.53	7.65	0.53
Drugs	Triamterene	Training	1.1	7.309	3.611	0.142	1.812	861	0.298	6	0.012	406	1.018	6.56	7.94	1.38	7.31	0.75
Drugs	Trichlormethiazide	Training	0.55	0.133	0.285	0.253	1.889	622	0.089	4	0.000	580	1.147	4.92	3.67	-1.25	4.32	-0.60
Explosives	1-MNG	Select	-0.43	0.684	0.000	0.172	0.000	0	0.000	2	0.000	1	0.000	1.85	3.10	1.25	3.85	2.00
Explosives	2,4-DNT	Select	2.1	5.014	0.000	0.113	0.000	14	0.000	0	0.000	58	0.776	11.21	10.15	-1.06	9.55	-1.66
Explosives	2,6-DA-4-NT	Select	0.92	1.395	0.000	0.086	0.000	8	0.000	4	0.000	44	3.178	3.16	5.83	2.67	6.33	3.17
Explosives	2-NT	Select	2.45	5.02	0.000	0.138	0.000	0	0.000	0	0.000	16	0.000	10.29	10.80	0.51	9.72	-0.57
Explosives	3,5-DNA	Select	1.63	1.945	0.000	0.069	0.000	24	0.000	2	0.000	62	0.000	9.26	8.85	-0.41	8.41	-0.85
Explosives	ETN	Select	3.21	1.266	1.525	0.068	1.332	114	0.000	0	0.000	166	0.794	11.60	10.52	-1.08	11.36	-0.24
Drugs	Andarine	Select	3.26	5.987	10.773	0.088	3.005	2276	0.799	3	0.009	1008	0.936	12.40	12.42	0.02	12.41	0.01
Drugs	Bambuterol	Select	-0.93	1.064	33.943	0.099	3.047	796	0.761	2	0.010	491	0.694	8.41	9.65	1.24	9.38	0.97
Drugs	Carphedon	Select	0.31	0.701	3.094	0.172	1.425	244	0.047	2	0.000	165	1.355	5.16	4.84	-0.32	5.86	0.70
Drugs	Desonide	Select	2.79	3.264	18.304	0.108	3.044	7390	0.993	2	0.010	3439	1.079	12.50	13.56	1.06	12.93	0.43
Drugs	Etamivan	Select	1.44	2.934	1.173	0.114	1.363	130	0.193	1	0.000	111	0.880	5.75	7.62	1.87	8.52	2.77
Drugs	Fencamfamin	Select	0.39	12.371	0.000	0.177	0.693	425	0.353	1	0.000	259	0.727	10.22	9.61	-0.61	8.23	-1.99
Drugs	Fludrocortisone	Select	1.54	2.987	12.545	0.123	2.583	4441	0.945	3	0.011	2227	0.877	9.59	9.70	0.11	9.68	0.09
Drugs	Fluticasone prop.	Select	3.71	11.64	23.138	0.088	3.379	8144	1.007	1	0.007	3998	1.135	16.21	16.34	0.13	16.11	-0.10
Drugs	Nadoxolol	Select	1.63	3.898	4.570	0.111	2.314	468	0.389	4	0.009	217	0.790	9.26	8.79	-0.48	8.57	-0.69
Drugs	Ostarine	Select	3.26	5.987	10.773	0.088	3.005	2276	0.799	3	0.009	1008	0.936	13.34	12.42	-0.92	12.41	-0.93
Drugs	Pemoline	Select	0.5	1.391	0.000	0.198	0.000	106	0.000	2	0.000	95	1.163	3.15	4.33	1.18	4.75	1.60
Drugs	Phendimetrazine	Select	1.28	3.404	0.887	0.153	0.000	170	0.275	0	0.000	92	0.214	6.32	7.31	0.99	7.36	1.04
Drugs	Propylhexedrine	Select	-0.3	6.262	0.000	0.158	0.000	16	0.039	1	0.000	17	1.375	8.76	6.37	-2.39	5.94	-2.83
Drugs	Strychnine	Select	0.21	8.431	1.484	0.168	1.609	7129	0.817	0	0.000	2406	1.691	8.14	9.69	1.55	8.80	0.66
Drugs	Terbutaline	Select	-1.94	1.275	9.263	0.147	1.906	116	0.299	4	0.000	141	0.205	2.83	2.70	-0.13	3.92	1.09
Drugs	Triamcinolone	Select	0.92	0.75	11.668	0.121	2.649	5110	0.925	4	0.009	2574	0.884	7.24	7.96	0.72	8.38	1.14
Explosives	1,2-DNG	Test	0.86	0.027	0.000	0.124	0.000	4	0.000	1	0.000	20	0.421	4.42	5.26	0.84	6.19	1.77

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	2,6-DNT	Test	2.1	5.014	0.000	0.105	0.000	12	0.000	0	0.000	68	0.000	11.04	10.92	-0.12	9.82	-1.22
Explosives	PYX	Test	7.04	18.199	8.482	0.104	3.682	8868	1.152	2	0.007	3148	0.659	16.90	17.68	0.78	17.90	1.00
Explosives	TATB	Test	2.93	0.01	0.000	0.068	0.000	150	0.000	6	0.000	384	0.235	9.40	8.94	-0.46	8.39	-1.01
Explosives	TATP	Test	2.16	3.13	0.000	0.111	0.000	120	0.246	0	0.000	96	0.000	11.99	9.38	-2.61	9.00	-3.00
Drugs	6-MAM	Test	0.5	5.473	5.070	0.155	2.127	3581	0.741	1	0.000	1650	1.406	4.45	7.49	3.04	8.42	3.97
Drugs	Aminoglutethimide	Test	1.27	1.896	0.134	0.152	0.732	345	0.212	3	0.000	229	1.692	4.99	5.75	0.76	6.42	1.43
Drugs	Atenolol	Test	-2.19	0.855	6.837	0.115	2.412	258	0.669	4	0.009	130	1.633	2.73	4.63	1.90	3.52	0.79
Drugs	Bendroflumethiazide	Test	1.48	2.503	11.010	0.153	2.775	2293	0.762	4	0.010	1289	0.841	8.50	7.98	-0.52	8.45	-0.05
Drugs	Bumetanide	Test	0.21	3.108	14.148	0.115	2.967	1105	0.761	4	0.009	598	0.932	6.25	7.83	1.58	7.99	1.74
Drugs	Bupropion	Test	2.69	10.356	1.577	0.182	2.043	113	0.452	1	0.000	158	1.229	9.54	10.80	1.26	10.98	1.44
Drugs	Chlorothiazide	Test	-0.1	0.035	0.000	0.229	0.000	270	0.000	3	0.000	341	0.821	2.08	2.24	0.16	3.00	0.92
Drugs	Clobetasol	Test	2.74	8.152	10.348	0.126	2.836	5110	0.966	2	0.009	2574	0.847	12.94	12.85	-0.09	12.46	-0.48
Drugs	Cyclopenthiazide	Test	1.51	1.396	7.940	0.192	2.538	1267	0.512	4	0.010	868	0.946	9.61	7.14	-2.47	7.22	-2.39
Drugs	Danazol	Test	4.33	17.257	8.778	0.137	2.713	4248	0.977	1	0.009	1647	0.983	16.65	16.39	-0.26	15.31	-1.34
Drugs	Gestrinone	Test	3.26	15.002	7.048	0.154	2.470	2558	0.847	1	0.007	1079	1.095	14.00	14.26	0.26	13.45	-0.55
Drugs	Labetalol	Test	0.53	7.151	12.048	0.116	2.800	1029	0.741	5	0.004	352	0.761	9.84	9.40	-0.45	9.04	-0.80
Drugs	Oxycodone	Test	0.15	1.907	2.814	0.145	0.732	3691	0.746	1	0.000	1757	1.292	4.60	4.68	0.08	6.11	1.51
Drugs	Phenmetrazine	Test	0.09	2.446	0.887	0.162	0.000	122	0.000	1	0.000	62	0.280	5.19	5.27	0.08	5.52	0.33
Drugs	Pindolol	Test	-0.65	1.707	5.411	0.123	2.266	357	0.583	3	0.011	178	1.745	4.86	6.54	1.68	5.38	0.52
Drugs	Sotalol	Test	-1.97	0.503	6.636	0.128	2.006	168	0.468	3	0.020	170	0.756	2.91	4.92	2.01	3.13	0.22
Drugs	Xipamide	Test	1.15	5.683	5.957	0.150	2.975	1108	0.559	4	0.012	639	0.583	5.01	7.72	2.71	8.12	3.11
OGSR	2,4-DNDPA	External	3.76	10.784	2.880	0.166	1.788	506	0.361	1	0.012	232	0.712	16.58	11.72	-4.86	13.87	-2.71
OGSR	2-NDPA	External	3.58	10.475	0.445	0.182	1.068	276	0.000	1	0.000	127	1.244	16.43	12.44	-3.99	14.37	-2.06
OGSR	4,4'-DNDPA	External	3.78	10.784	3.634	0.187	2.096	536	0.296	1	0.016	180	0.305	14.87	10.39	-4.48	13.97	-0.90
OGSR	4-NDPA	External	3.36	10.475	2.173	0.198	1.623	288	0.000	1	0.012	100	0.263	14.19	10.66	-3.53	14.22	0.03
OGSR	Dimethyl phthalate	External	1.88	4.018	0.000	0.086	0.000	56	0.000	0	0.000	70	1.517	7.38	9.29	1.91	7.50	0.12

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
OGSR	N-nitroso-DPA	External	3.03	10.385	1.719	0.191	0.693	220	0.000	0	0.000	92	1.134	13.45	12.31	-1.14	14.04	0.59
MEKPs	MEKP CP dimer	External	2.08	3.216	0.000	0.154	0.000	0	0.000	0	0.000	20	0.000	-	8.89	-	8.49	-
MEKPs	MEKP CP tetramer	External	4.50	11.535	8.758	0.074	2.565	876	0.997	0	0.000	328	1.125	-	17.10	-	15.49	-
MEKPs	MEKP CP trimer	External	3.44	6.996	1.915	0.094	0.000	267	0.445	0	0.000	138	0.708	-	13.64	-	11.77	-
MEKPs	MEKP DHP dimer	External	3.19	2.067	0.438	0.096	0.000	0	0.182	2	0.000	16	0.948	-	9.83	-	9.61	-
MEKPs	MEKP DHP monomer	External	1.33	0.093	0.000	0.127	0.000	0	0.000	2	0.000	0	0.000	-	6.13	-	6.54	-
MEKPs	MEKP DHP pentamer	External	8.28	14.812	34.125	0.055	3.325	1398	1.026	2	0.012	496	0.973	-	21.78	-	19.44	-
MEKPs	MEKP DHP tetetramer	External	6.65	9.800	17.611	0.066	2.937	672	1.002	2	0.011	264	0.962	-	18.50	-	16.75	-
MEKPs	MEKP DHP trimer	External	4.91	5.483	3.751	0.077	2.296	206	0.439	2	0.000	104	0.950	-	13.78	-	13.76	-
MNs	M-1,2,3,4,5-PN	External	3.74	1.968	1.462	0.048	2.149	551	0.676	1	0.000	560	0.929	-	10.50	-	11.83	-
MNs	M-1,2,3,4,6-PN	External	3.79	1.968	6.025	0.046	2.279	537	0.624	1	0.008	513	0.947	-	12.18	-	12.21	-
MNs	M-1,2,3,5,6-PN	External	3.74	1.968	4.411	0.057	2.277	554	0.670	1	0.006	501	1.000	-	11.60	-	11.89	-
MNs	M-1,2,3,4-TN	External	2.28	0.284	3.338	0.052	1.759	336	0.351	2	0.000	391	0.874	-	8.28	-	9.80	-
MNs	M-1,2,3,5-TN	External	2.35	0.284	3.612	0.055	1.889	353	0.303	2	0.000	383	0.935	-	8.46	-	9.98	-
MNs	M-1,2,4,5-TN	External	2.60	0.284	2.307	0.065	1.935	371	0.335	2	0.000	383	0.960	-	8.30	-	9.93	-
MNs	M-1,3,4,5-TN	External	2.57	0.284	1.311	0.052	1.824	356	0.316	2	0.000	392	0.895	-	8.35	-	9.93	-
MNs	M-2,3,4,5-TN	External	2.32	0.284	0.231	0.054	1.529	356	0.321	2	0.000	430	0.874	-	7.72	-	9.33	-
MNs	M-1,2,3,6-TN	External	2.32	0.284	3.331	0.057	1.943	349	0.304	2	0.008	343	0.960	-	9.31	-	9.59	-
MNs	M-1,2,4,6-TN	External	2.60	0.284	2.792	0.068	2.055	357	0.325	2	0.008	345	0.967	-	9.34	-	9.71	-
MNs	M-1,3,4,6-TN	External	2.73	0.284	5.823	0.049	2.096	352	0.314	2	0.017	354	0.921	-	9.22	-	9.86	-
MNs	M-1,2,5,6-TN	External	2.28	0.284	4.013	0.053	1.940	370	0.353	2	0.006	334	1.032	-	9.30	-	9.73	-
MNs	M-1,2,3-TriN	External	0.72	0.091	1.872	0.065	1.347	179	0.155	3	0.000	245	0.863	-	5.75	-	7.19	-
MNs	M-1,2,4-TriN	External	0.92	0.091	0.010	0.077	1.606	206	0.039	3	0.000	247	0.899	-	5.86	-	7.29	-
MNs	M-1,3,4-TriN	External	0.87	0.091	2.363	0.057	1.519	203	0.155	3	0.000	253	0.847	-	6.21	-	7.65	-
MNs	M-2,3,4-TriN	External	0.80	0.091	0.001	0.060	0.806	194	0.075	3	0.000	282	0.818	-	5.73	-	6.94	-
MNs	M-1,2,6-TriN	External	0.69	0.091	5.293	0.067	1.613	213	0.141	3	0.008	209	0.992	-	7.66	-	7.45	-

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Average	Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
MNs	M-1,3,6-TriN	External	0.91	0.091	4.309	0.066	1.820	214	0.166	3	0.017	217	0.947	-	7.29	-	7.12	-
SNs	S-2,3,4,5,6-PN	External	3.74	1.968	3.688	0.050	2.149	551	0.676	1	0.000	560	0.929	-	10.99	-	12.10	-
SNs	S-1,3,4,5,6-PN	External	3.79	1.968	6.332	0.048	2.279	537	0.624	1	0.008	513	0.947	-	12.21	-	12.22	-
SNs	S-1,2,4,5,6-PN	External	3.74	1.968	4.190	0.064	2.277	554	0.670	1	0.006	501	1.000	-	11.43	-	11.77	-
SNs	S-1,2,3,5,6-PN	External	3.74	1.968	6.749	0.055	2.277	554	0.670	1	0.006	501	1.000	-	12.12	-	12.21	-
SNs	S-1,2,3,4,6-PN	External	3.79	1.968	6.661	0.051	2.279	537	0.624	1	0.008	513	0.947	-	12.22	-	12.22	-
SNs	S-1,2,3,4,5-PN	External	3.74	1.968	2.746	0.052	2.149	551	0.676	1	0.000	560	0.929	-	10.72	-	11.95	-
SNs	S-3,4,5,6-TN	External	2.28	0.284	3.555	0.059	1.759	336	0.351	2	0.000	391	0.874	-	8.20	-	9.73	-
SNs	S-2,4,5,6-TN	External	2.35	0.284	4.190	0.065	1.889	353	0.303	2	0.000	383	0.935	-	8.40	-	9.91	-
SNs	S-2,3,5,6-TN	External	2.60	0.284	2.719	0.073	1.935	371	0.335	2	0.000	383	0.960	-	8.25	-	9.87	-
SNs	S-2,3,4,6-TN	External	2.57	0.284	3.741	0.058	1.824	356	0.316	2	0.000	392	0.895	-	8.76	-	10.17	-
SNs	S-2,3,4,5-TN	External	2.32	0.284	0.000	0.058	1.529	356	0.321	2	0.000	430	0.874	-	7.60	-	9.24	-
SNs	S-1,4,5,6-TN	External	2.32	0.284	4.625	0.060	1.943	349	0.304	2	0.008	343	0.960	-	9.46	-	9.70	-
SNs	S-1,3,5,6-TN	External	2.60	0.284	1.998	0.071	2.055	357	0.325	2	0.008	345	0.967	-	9.19	-	9.57	-
SNs	S-1,3,4,6-TN	External	2.73	0.284	6.238	0.057	2.096	352	0.314	2	0.017	354	0.921	-	9.18	-	9.80	-
SNs	S-1,3,4,5-TN	External	2.57	0.284	3.655	0.058	1.824	356	0.316	2	0.000	392	0.895	-	8.74	-	10.16	-
SNs	S-1,2,5,6-TN	External	2.28	0.284	2.842	0.063	1.940	370	0.353	2	0.006	334	1.032	-	8.97	-	9.44	-
SNs	S-1,2,4,6-TN	External	2.60	0.284	6.084	0.060	2.055	357	0.325	2	0.008	345	0.967	-	9.95	-	10.22	-
SNs	S-1,2,4,5-TN	External	2.60	0.284	3.570	0.063	1.935	371	0.335	2	0.000	383	0.960	-	8.60	-	10.13	-
SNs	S-1,2,3,6-TN	External	2.32	0.284	4.941	0.056	1.943	349	0.304	2	0.008	343	0.960	-	9.56	-	9.80	-
SNs	S-1,2,3,5-TN	External	2.35	0.284	4.661	0.057	1.889	353	0.303	2	0.000	383	0.935	-	8.64	-	10.09	-
SNs	S-1,2,3,4-TN	External	2.28	0.284	2.871	0.055	1.759	336	0.351	2	0.000	391	0.874	-	8.13	-	9.69	-
SNs	S-1,2,3-TrN	External	0.72	0.091	1.424	0.064	1.347	179	0.155	3	0.000	245	0.863	-	5.69	-	7.14	-
SNs	S-1,2,4-TrN	External	0.92	0.091	2.785	0.069	1.606	206	0.039	3	0.000	247	0.899	-	6.43	-	7.77	-
SNs	S-1,2,5-TrN	External	0.87	0.091	4.614	0.068	1.601	220	0.020	3	0.000	240	0.992	-	6.72	-	7.98	-
SNs	S-1,2,6-TrN	External	0.69	0.091	5.407	0.074	1.613	213	0.141	3	0.008	209	0.992	-	7.60	-	7.37	-

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
SNs	S-1,3,4-TrN	External	0.87	0.091	2.848	0.066	1.519	203	0.155	3	0.000	253	0.847	-	6.16	-	7.58	-
SNs	S-1,3,5-TrN	External	1.03	0.091	3.204	0.066	1.622	212	0.068	3	0.000	248	0.909	-	6.63	-	7.98	-
SNs	S-1,3,6-TrN	External	0.69	0.091	5.318	0.073	1.613	213	0.141	3	0.008	209	0.992	-	7.60	-	7.37	-
SNs	S-2,3,4-TrN	External	0.80	0.091	0.000	0.066	0.806	194	0.075	3	0.000	282	0.818	-	5.63	-	6.85	-
SNs	S-2,3,5-TrN	External	1.05	0.091	0.173	0.071	1.332	212	0.026	3	0.000	277	0.899	-	6.11	-	7.43	-
SNs	S-2,3,6-TrN	External	0.83	0.091	3.805	0.068	1.514	219	0.146	3	0.000	246	0.932	-	6.26	-	7.63	-
SNs	S-2,4,5-TrN	External	1.05	0.091	0.000	0.071	1.332	212	0.026	3	0.000	277	0.899	-	6.08	-	7.41	-
SNs	S-2,4,6-TrN	External	1.03	0.091	3.168	0.071	1.622	212	0.068	3	0.000	248	0.909	-	6.56	-	7.90	-
SNs	S-2,5,6-TrN	External	0.87	0.091	5.148	0.072	1.601	220	0.020	3	0.000	240	0.992	-	6.75	-	7.99	-
SNs	S-3,4,5-TrN	External	0.80	0.091	0.010	0.067	0.806	194	0.075	3	0.000	282	0.818	-	5.62	-	6.83	-
SNs	S-3,4,6-TrN	External	0.87	0.091	3.357	0.066	1.519	203	0.155	3	0.000	253	0.847	-	6.25	-	7.65	-
SNs	S-3,5,6-TrN	External	0.92	0.091	0.061	0.080	1.606	206	0.039	3	0.000	247	0.899	-	5.83	-	7.25	-
SNs	S-4,5,6-TrN	External	0.72	0.091	2.129	0.068	1.347	179	0.155	3	0.000	245	0.863	-	5.76	-	7.18	-
SNs	S-1,2-DN	External	-0.70	1.198	1.774	0.084	1.105	90	0.000	4	0.000	135	0.900	-	4.46	-	5.25	-
SNs	S-1,3-DN	External	-0.46	1.198	2.100	0.078	1.233	96	0.001	4	0.000	140	0.844	-	4.93	-	5.74	-
SNs	S-1,4-DN	External	-0.46	1.198	2.771	0.090	1.320	112	0.004	4	0.000	140	0.893	-	4.91	-	5.70	-
SNs	S-1,5-DN	External	-0.52	1.198	4.629	0.080	1.227	116	0.016	4	0.000	136	0.985	-	5.18	-	5.94	-
SNs	S-1,6-DN	External	-0.69	1.198	3.861	0.102	1.438	114	0.109	4	0.017	114	0.985	-	6.05	-	4.62	-
SNs	S-2,3-DN	External	-0.54	1.198	0.341	0.078	0.346	88	0.000	4	0.000	160	0.830	-	4.32	-	4.95	-
SNs	S-2,4-DN	External	-0.25	1.198	0.000	0.081	0.806	100	0.000	4	0.000	162	0.844	-	4.74	-	5.46	-
SNs	S-2,5-DN	External	-0.21	1.198	0.161	0.083	1.086	112	0.005	4	0.000	158	0.985	-	4.85	-	5.64	-
SNs	S-2,6-DN	External	-0.52	1.198	3.891	0.085	1.227	116	0.016	4	0.000	136	0.985	-	5.01	-	5.77	-
SNs	S-3,4-DN	External	-0.60	1.198	0.032	0.077	0.000	90	0.000	4	0.000	164	0.766	-	4.13	-	4.68	-
SNs	S-3,5-DN	External	-0.25	1.198	0.016	0.077	0.806	100	0.000	4	0.000	162	0.844	-	4.80	-	5.52	-
SNs	S-3,6-DN	External	-0.46	1.198	2.117	0.079	1.320	112	0.004	4	0.000	140	0.893	-	4.94	-	5.77	-
SNs	S-4,5-DN	External	-0.54	1.198	0.016	0.090	0.346	88	0.000	4	0.000	160	0.830	-	4.11	-	4.73	-

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors										Average	Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
SNs	S-4,6-DN	External	-0.46	1.198	3.275	0.091	1.233	96	0.001	4	0.000	140	0.844	-	4.95	-	5.70	-
SNs	S-5,6-DN	External	-0.70	1.198	2.118	0.083	1.105	90	0.000	4	0.000	135	0.900	-	4.52	-	5.31	-
SNs	S-1-MN	External	-2.18	3.361	1.772	0.119	0.958	40	0.000	5	0.000	61	0.918	-	3.32	-	3.28	-
SNs	S-2-MN	External	-1.79	3.361	0.367	0.098	0.346	32	0.000	5	0.000	74	0.918	-	3.59	-	3.54	-
SNs	S-3-MN	External	-1.98	3.361	0.000	0.091	0.000	30	0.000	5	0.000	75	0.788	-	3.24	-	3.21	-
SNs	S-4-MN	External	-1.98	3.361	0.078	0.105	0.000	30	0.000	5	0.000	75	0.788	-	3.08	-	3.03	-
SNs	S-5-MN	External	-1.79	3.361	0.168	0.093	0.346	32	0.000	5	0.000	74	0.918	-	3.62	-	3.58	-
SNs	S-6-MN	External	-2.18	3.361	2.089	0.119	0.958	40	0.000	5	0.000	61	0.918	-	3.35	-	3.32	-
$C_3H_6O_6N_6$	2,4,6-Trinitro-1,3,5-triazinane	External	-3.18	1.963	0.000	0.183	0.000	72	0.000	3	0.000	96	0.000	-	-0.30	-	0.78	-
$C_3H_6O_6N_6$	N,N',N''-Trinitro-1,2,3-cyclopropanetriamine	External	-4.56	0.034	0.000	0.132	0.000	30	0.000	3	0.000	108	0.830	-	-2.67	-	-0.47	-
$C_3H_6O_6N_6$	1,3,5-Trinitro-1,2,3-triazinane	External	-1.27	0.990	0.000	0.170	0.000	72	0.000	1	0.000	96	0.000	-	2.05	-	3.24	-
$C_3H_6O_6N_6$	2-(N'-Nitro-N-nitroso-carbamimidamido)ethyl nitrate	External	-2.14	0.065	0.000	0.130	0.759	25	0.000	2	0.000	56	1.389	-	1.25	-	2.65	-
$C_3H_6O_6N_6$	1,2,2-Trinitro-1,3,5-triazinane	External	0.63	3.265	0.000	0.184	0.000	36	0.000	2	0.000	157	0.000	-	5.91	-	5.69	-
$C_{17}H_7O_{16}N_{11}$	3,6-Dinitro-N,N'-bis(2,4,6- trinitrophenyl)-2,5-pyridinediamine	External	5.78	18.199	19.868	0.103	3.572	8860	1.179	2	0.009	3148	0.841	-	17.28	-	17.66	-
$C_{17}H_7O_{16}N_{11}$	2,6-Dinitro-N,N'-bis(2,4,6- trinitrophenyl)-3,5-pyridinediamine	External	6.09	18.199	19.716	0.095	3.687	8868	1.152	2	0.007	3148	0.801	-	17.59	-	18.01	-
$C_6H_6O_6N_6$	4,5,6-Trinitro-1,2,3-benzenetriamine	External	2.48	0.010	0.000	0.067	0.000	214	0.000	6	0.000	360	0.423	-	8.20	-	7.83	-
$C_6H_6O_6N_6$	(1Z,2Z,3E,4Z,5Z,6E)-N,N',N'',N''',N'''',N''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	External	-4.96	3.884	1.248	0.055	0.000	234	0.000	6	0.000	264	0.352	-	-0.38	-	0.61	-
$C_6H_6O_6N_6$	N,N',N''-Trihydroxy-2,4,6-trinitroso-1,3,5-benzenetriamine	External	-2.56	2.492	0.000	0.097	0.000	234	0.000	6	0.000	264	0.352	-	1.64	-	2.03	-

Table A. 5 (Continued): Molecular descriptors and predicted retention times for best network and ensemble following genetic algorithm feature selection.

				Molecular descriptors											Best ANN, MLP:11-3-1		Ensemble, 6xMLP:11-3-1	
Dataset	Analyte	ANN subset	LogD (pH=7)	MLOGP2	RDF090e	HATS3p	ATS8p	p3p5	BELe8	H-050	JGI10	p1p4B	GATS6m	Average t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
$C_6H_6O_6N_6$	N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	External	-4.96	3.884	1.248	0.055	0.000	234	0.000	6	0.000	264	0.352	-	-0.38	-	0.61	-
$C_6H_6O_6N_6$	1-[(2Z)-2-Methyl-3-nitro-2-propen-1-yl]-3,5-dinitro-1H-1,2,4-triazole	External	0.28	5.866	0.000	0.144	1.086	157	0.043	0	0.000	234	1.380	-	7.15	-	7.59	-
$C_6H_6O_6N_6$	1-(2-Methyl-3-nitro-2-propen-1-yl)-3,5-dinitro-1H-1,2,4-triazole	External	0.28	5.866	2.233	0.134	1.086	157	0.043	0	0.000	234	1.380	-	7.61	-	8.01	-
$C_6H_6O_6N_6$	(1E,2E,3E,4E,5E,6E)-N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-cyclohexanehexaimine	External	-4.96	3.884	1.248	0.055	0.000	234	0.000	6	0.000	264	0.352	-	-0.38	-	0.61	-
$C_6H_6O_6N_6$	N-{1-[Amino(oxo)acetyl]-3,5-dioxo-1,2,4-triazolidin-4-yl}ethanediamide	External	-6.27	6.284	2.952	0.147	1.421	226	0.000	6	0.000	246	0.619	-	-0.34	-	-0.34	-

Table A. 6: Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																	
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4-3N	IDE	nR08
Explosives	1,2-DNB	Training	1.60	0.000	16	-3.11	0	0.284	73.9	9.055	0.027	0.497	2.120	4	3.949	0.00	42.3	24	0	2.28	0
Explosives	1,2-DNG	Training	0.86	0.936	13	-1.16	0	1.037	150.2	4.326	1.330	0.625	2.206	0	3.680	0.09	66.0	12	0	2.73	0
Explosives	1,3-DNB	Training	1.55	0.602	16	-3.11	0	0.284	76.3	9.005	0.748	0.497	0.197	0	3.934	0.00	44.2	18	0	2.47	0
Explosives	1-MNG	Training	-0.43	0.346	9	-0.53	0	0.992	85.5	6.549	0.870	0.679	0.051	0	3.619	0.00	36.0	1	0	2.46	0
Explosives	2,4-DA-6-NT	Training	0.92	0.000	17	-2.44	0	0.331	66.2	10.639	0.138	0.676	1.371	4	3.930	0.00	41.8	12	0	2.26	0
Explosives	2,4-DNT	Training	2.10	1.006	18	-3.45	0	0.268	84.0	9.796	0.822	0.610	1.655	4	3.955	0.00	52.5	27	0	2.50	0
Explosives	2-Am-4,6-DNT	Training	1.88	1.006	20	-2.98	0	0.313	93.5	14.405	0.827	0.661	1.235	8	3.972	0.00	61.0	30	0	2.50	0
Explosives	2-NT	Training	2.45	0.000	13	-3.45	0	0.207	47.3	5.808	0.014	0.616	2.171	0	3.892	0.00	27.7	8	0	2.15	0
Explosives	3,4-DNT	Training	2.15	0.647	18	-3.45	0	0.268	82.5	10.825	0.362	0.610	3.352	6	3.965	0.00	51.3	29	0	2.41	0
Explosives	3,5-DNA	Training	1.63	0.602	18	-2.65	0	0.333	86.1	14.740	0.764	0.584	0.127	0	3.950	0.00	52.3	20	0	2.48	0
Explosives	4-NT	Training	2.45	0.647	13	-3.45	0	0.207	49.3	4.602	0.389	0.616	1.214	0	3.883	0.00	29.3	6	0	2.34	0
Explosives	DADP	Training	1.23	0.000	16	-2.37	0	0.750	95.5	4.870	0.037	0.378	0.233	0	3.652	0.49	28.7	0	0	2.13	0
Explosives	DEDPU	Training	3.26	3.098	27	-4.83	1	0.871	183.5	23.795	5.455	0.471	15.504	352	3.866	0.66	136.1	94	16	3.05	0
Explosives	DMNB	Training	1.82	0.000	18	-2.93	0	0.310	80.0	5.865	0.007	0.412	1.297	0	3.874	0.00	66.0	8	0	2.16	0
Explosives	EGDN	Training	1.51	0.759	10	-1.37	0	1.092	123.7	3.589	1.342	0.547	2.108	0	3.643	0.00	45.0	6	0	2.73	0
Explosives	HMDD	Training	0.00	0.000	16	-1.66	0	0.964	111.9	2.572	0.002	0.422	0.235	27	3.649	0.05	21.3	34	6	2.19	2
Explosives	HMTA	Training	0.98	0.000	18	-2.37	0	1.875	56.8	0.700	0.000	0.313	0.000	6	3.820	0.00	13.4	96	12	1.81	3
Explosives	HMTD	Training	0.13	0.000	18	-1.77	0	1.084	171.6	6.473	0.003	0.373	3.860	69	3.612	0.25	29.4	48	6	2.29	0
Explosives	HMX	Training	-1.71	1.996	28	-3.56	1	0.679	261.1	14.552	1.641	0.432	18.086	114	3.855	0.04	122.6	272	20	2.85	1
Explosives	HNDPA	Training	4.44	3.407	47	-4.30	1	0.390	370.0	43.669	7.469	0.421	11.694	1297	4.065	0.36	303.6	470	4	3.33	0
Explosives	MHN	Training	5.16	3.115	38	-3.51	1	1.092	571.0	38.176	7.246	0.428	17.963	357	3.818	0.86	435.0	214	0	3.24	0
Explosives	NB	Training	1.81	0.000	11	-3.11	0	0.214	41.0	3.511	0.015	0.472	0.146	0	3.862	0.00	22.4	4	0	2.15	0
Explosives	NG	Training	2.41	1.502	17	-1.87	0	1.092	225.5	15.971	4.354	0.515	1.020	0	3.708	0.08	105.0	28	0	2.93	0
Explosives	NM	Training	0.10	0.000	3	-0.91	0	0.545	14.3	0.000	0.000	0.614	0.000	0	3.558	0.00	6.0	0	0	1.00	0
Explosives	PA	Training	-1.40	1.359	23	-2.79	0	0.471	125.5	19.405	1.308	0.566	0.614	24	4.005	0.00	82.2	64	0	2.56	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
Explosives	PGDN	Training	1.87	0.759	12	-1.87	0	0.854	133.7	4.582	1.394	0.571	1.974	0	3.671	0.00	55.0	10	0	2.71	0
Explosives	R-Salt	Training	-1.18	0.482	15	-0.72	0	0.689	116.4	2.074	0.088	0.513	7.332	0	3.679	0.00	42.9	63	9	2.46	0
Explosives	SHN	Training	5.16	3.115	38	-3.51	1	1.092	571.0	46.694	7.524	0.428	30.904	357	3.818	0.83	435.0	214	0	3.24	0
Explosives	TATP	Training	2.16	0.000	24	-3.00	0	0.778	194.8	22.141	0.194	0.334	29.694	12	3.659	0.84	55.9	0	0	2.30	0
Explosives	TEGDN	Training	1.46	1.869	16	-1.65	0	1.090	305.4	10.127	7.377	0.448	6.666	26	3.621	0.00	120.0	18	0	3.57	0
Explosives	Tetryl	Training	1.25	2.031	29	-3.84	0	0.413	191.3	22.313	2.852	0.546	4.283	92	4.030	0.00	134.9	156	4	2.77	0
Explosives	TMETN	Training	3.08	1.537	21	-2.69	0	0.803	254.4	26.192	4.252	0.496	0.586	0	3.737	0.00	136.0	24	0	2.96	0
Explosives	TNB	Training	1.36	1.247	21	-3.23	0	0.349	113.8	18.009	1.293	0.474	0.153	0	3.988	0.00	72.5	42	0	2.56	0
Explosives	TNT	Training	1.79	1.479	23	-3.56	0	0.326	121.0	20.966	1.323	0.567	1.221	24	4.009	0.00	82.2	64	0	2.56	0
Drugs	Acebutolol	Training	-0.73	3.017	31	-2.83	1	0.832	329.0	33.468	12.370	0.594	10.019	277	3.888	0.82	225.5	40	6	3.63	0
Drugs	Acetazolamide	Training	-0.58	1.595	19	0.18	0	1.626	88.3	1.217	3.690	0.766	2.388	6	4.684	0.01	58.9	30	8	2.68	0
Drugs	AICAR	Training	-2.49	2.220	28	0.01	0	1.059	176.2	8.344	5.186	0.724	13.398	291	3.953	0.37	96.3	91	10	2.92	0
Drugs	Alprenolol	Training	0.34	2.528	22	-3.57	1	0.869	200.3	24.095	7.653	0.607	9.034	95	3.840	0.80	117.0	10	2	3.28	0
Drugs	Altizide	Training	1.08	3.006	35	-1.86	1	1.328	195.5	15.906	6.336	0.713	11.774	390	4.684	0.18	129.9	60	10	3.26	0
Drugs	Aminoglutethimide	Training	1.27	2.232	26	-2.63	1	0.621	112.5	20.408	3.935	0.637	9.256	155	3.904	0.08	82.5	16	4	2.81	0
Drugs	Anastrazole	Training	2.68	3.244	34	-3.98	1	0.827	178.1	33.579	5.059	0.563	14.840	308	3.954	0.54	158.1	112	6	3.05	0
Drugs	Bambuterol	Training	-0.93	3.311	37	-2.30	1	1.050	370.7	29.952	8.520	0.541	22.770	254	3.886	0.74	257.9	53	3	3.40	0
Drugs	Bendroflumethiazide	Training	1.48	3.273	46	-2.82	1	1.539	279.6	24.340	6.824	0.634	18.878	1133	4.681	0.23	189.6	86	10	3.36	0
Drugs	Benzoylecgonine	Training	-0.21	2.898	33	-3.38	1	0.761	217.9	21.512	6.282	0.598	7.717	640	3.878	0.39	114.9	71	4	3.10	0
Drugs	Benzthiazide	Training	0.99	3.197	42	-3.02	1	1.227	258.4	18.861	10.626	0.641	15.321	964	4.687	0.24	182.5	82	10	3.57	0
Drugs	Beta\dexameth.	Training	1.92	3.337	53	-3.08	1	0.546	264.3	47.182	8.449	0.573	36.423	2577	4.013	1.18	128.5	0	0	3.15	0
Drugs	Budesonide	Training	3.02	3.644	57	-3.23	1	0.709	352.9	61.752	9.164	0.529	38.028	4022	4.018	1.21	144.9	0	0	3.37	1
Drugs	Bumetanide	Training	0.21	3.495	36	-2.99	1	1.473	268.3	34.113	6.788	0.648	15.618	460	4.623	0.55	199.9	36	2	3.20	0
Drugs	Bumetanide	Training	0.21	3.495	36	-2.99	1	1.473	268.3	34.113	6.788	0.648	15.618	460	4.623	0.55	199.9	36	2	3.20	0
Drugs	Bunolol	Training	0.15	2.598	31	-3.53	1	0.767	246.8	34.912	9.722	0.580	4.684	277	3.920	0.84	129.4	13	3	3.35	0
Drugs	Bupropion	Training	2.69	2.537	23	-4.38	1	0.522	118.6	17.561	4.982	0.565	9.698	75	3.946	0.14	93.8	6	6	2.89	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
Drugs	Butizide	Training	0.78	2.886	35	-1.93	1	1.282	172.4	15.322	5.350	0.657	18.268	370	4.684	0.29	113.4	54	10	3.06	0
Drugs	Canrenone	Training	2.50	3.208	48	-5.27	1	0.627	227.9	42.667	8.839	0.500	25.027	2238	3.996	0.73	92.4	0	0	3.19	0
Drugs	Carteolol	Training	-0.58	2.585	31	-2.56	1	0.830	254.8	32.295	9.823	0.606	6.083	265	3.907	0.87	131.2	40	7	3.39	0
Drugs	Celiprolol	Training	-0.41	3.183	37	-2.93	1	0.896	412.1	41.331	12.456	0.543	13.864	424	3.891	1.12	294.8	73	12	3.73	0
Drugs	Chlorexolone	Training	1.49	2.896	35	-3.19	1	0.944	177.8	15.463	7.637	0.627	9.350	599	4.629	0.19	106.7	83	8	3.10	0
Drugs	Clobenzorex	Training	2.89	2.779	24	-5.51	1	0.482	153.9	14.334	3.872	0.528	16.351	220	3.945	0.36	102.6	16	2	3.26	0
Drugs	Clobetasol	Training	2.74	3.391	53	-4.03	1	0.493	258.4	50.311	8.669	0.580	35.998	2577	4.017	0.96	128.5	0	0	3.15	0
Drugs	Clopamide	Training	1.75	2.950	34	-3.23	1	0.991	225.3	22.949	7.914	0.628	10.908	495	4.628	0.33	150.0	100	12	3.20	0
Drugs	Cyclopenthiazide	Training	1.51	3.145	39	-2.44	1	1.208	209.1	26.096	4.745	0.609	20.653	778	4.684	0.33	131.1	68	10	3.24	0
Drugs	Danazol	Training	4.33	3.284	48	-5.27	0	0.517	212.7	45.547	7.194	0.550	30.828	2217	4.017	0.72	82.9	44	1	3.15	0
Drugs	Diacetolol	Training	-1.40	2.799	29	-2.37	1	0.852	280.9	28.332	11.077	0.619	7.242	200	3.886	0.96	184.4	28	6	3.48	0
Drugs	Dichlorphenamide	Training	0.89	1.659	26	-1.51	0	1.335	90.7	6.297	1.841	0.637	6.244	15	4.675	0.32	83.5	0	0	2.54	0
Drugs	Efaproxiral	Training	0.38	3.031	37	-4.29	1	0.738	322.8	25.554	10.333	0.588	15.319	571	3.885	0.45	228.4	25	4	3.68	0
Drugs	Epitizide	Training	1.01	3.003	40	-2.06	1	1.404	262.9	14.261	7.629	0.722	13.472	480	4.684	0.18	164.9	69	10	3.39	0
Drugs	Etacr. acid	Training	-0.24	2.745	26	-4.18	1	0.585	180.1	15.224	6.653	0.730	9.370	108	4.071	0.31	131.7	0	0	3.18	0
Drugs	Etamivan	Training	1.44	2.249	21	-2.95	1	0.897	134.0	22.145	2.998	0.598	5.697	64	3.873	0.49	89.9	14	5	2.85	0
Drugs	Famprofazone	Training	1.18	3.614	42	-5.49	0	0.811	342.5	40.987	10.676	0.504	45.302	1262	3.930	1.34	255.0	329	18	3.52	0
Drugs	Fenbutrazate	Training	4.45	3.193	38	-4.47	1	0.922	379.4	45.291	6.840	0.504	44.644	1038	3.863	0.59	238.2	82	5	3.68	0
Drugs	Fencamfamin	Training	0.39	2.453	25	-4.66	1	0.889	105.6	19.482	0.424	0.522	13.497	243	3.927	0.44	59.2	19	2	2.72	0
Drugs	Fludrocortisone	Training	1.54	3.233	51	-2.96	1	0.548	254.5	49.928	8.014	0.540	37.399	2192	3.999	1.03	118.8	0	0	3.15	0
Drugs	Flumethasone	Training	1.74	3.386	55	-3.18	1	0.564	283.5	50.836	8.086	0.584	40.334	2923	4.018	1.27	138.6	0	0	3.15	0
Drugs	Formoterol	Training	-0.41	2.974	34	-2.90	1	0.792	321.0	28.148	7.711	0.631	23.274	570	3.880	0.45	226.5	54	4	3.73	0
Drugs	FPCAM	Training	-1.11	3.653	59	-4.31	1	0.625	365.0	68.119	7.498	0.574	40.832	3725	4.021	1.33	187.7	0	0	3.28	0
Drugs	Fulvestrant	Training	7.56	3.846	65	-7.09	0	1.041	786.6	81.092	19.933	0.444	41.891	2788	4.233	1.50	496.2	0	0	4.29	0
Drugs	Furosemide	Training	-0.74	2.870	32	-1.73	1	1.031	190.0	18.808	7.548	0.765	6.606	373	4.628	0.12	143.5	34	2	3.17	0
Drugs	Hydroflumeth.	Training	0.23	2.305	35	-1.25	1	1.756	158.0	14.916	2.737	0.717	5.034	265	4.679	0.22	95.9	42	5	2.73	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
Drugs	Labetalol	Training	0.53	2.954	33	-3.86	1	0.609	280.0	29.140	11.853	0.622	19.616	482	3.887	0.62	206.7	26	2	3.65	0
Drugs	Mefruside	Training	1.67	3.058	38	-1.80	1	1.273	222.3	30.871	4.565	0.623	23.666	552	4.681	0.23	183.9	73	3	3.20	0
Drugs	Metoprolol	Training	-0.59	2.483	23	-2.89	1	1.000	247.8	23.058	11.766	0.571	11.051	106	3.832	0.80	143.6	11	2	3.56	0
Drugs	Nadoxolol	Training	1.63	2.551	26	-3.20	1	0.635	199.9	20.815	7.662	0.648	2.478	223	4.008	0.18	103.2	6	0	3.24	0
Drugs	Nikethamide	Training	0.54	1.833	16	-2.29	0	0.872	85.6	10.809	0.952	0.592	6.774	24	3.814	0.43	58.5	17	6	2.62	0
Drugs	Ostarine	Training	3.26	3.156	47	-3.64	1	0.699	505.3	30.983	14.125	0.664	7.121	1039	3.947	0.76	367.6	152	8	3.89	0
Drugs	Oxprenolol	Training	-0.15	2.652	23	-3.06	1	1.000	244.6	26.333	8.746	0.601	6.655	106	3.829	0.72	130.9	11	2	3.36	0
Drugs	Oxycodone	Training	0.15	2.857	45	-2.63	1	0.816	190.1	41.042	3.104	0.603	21.776	1700	4.078	0.17	62.2	88	2	2.72	1
Drugs	Oxymorphone	Training	0.34	2.596	44	-2.41	1	0.698	165.4	36.690	1.752	0.639	22.202	1499	4.076	0.19	53.8	83	2	2.60	1
Drugs	Pemoline	Training	0.50	1.465	19	-2.44	0	0.889	84.4	6.574	0.723	0.671	4.496	69	3.862	0.09	42.5	8	4	2.56	0
Drugs	Pentazocine	Training	1.89	2.944	35	-4.91	1	0.667	175.3	41.273	6.322	0.547	24.946	481	3.980	0.85	88.1	69	5	3.10	1
Drugs	Pentetrazol	Training	0.22	0.000	14	-2.90	0	0.511	50.2	0.543	0.006	0.532	0.133	9	3.871	0.00	14.8	37	10	2.05	0
Drugs	Phendimetrazine	Training	1.28	2.031	20	-3.07	0	1.094	92.4	12.686	1.359	0.549	10.294	75	3.843	0.26	48.2	20	2	2.61	0
Drugs	Polythiazide	Training	1.59	3.169	42	-2.32	1	1.382	274.4	16.860	7.592	0.705	15.396	557	4.689	0.25	177.2	115	10	3.37	0
Drugs	Prenylamine	Training	2.60	3.337	34	-6.52	0	0.926	266.0	37.383	7.175	0.439	29.111	709	3.897	0.40	210.9	19	2	3.55	0
Drugs	Probenecid	Training	0.13	2.710	26	-3.20	1	1.545	164.9	24.936	5.754	0.589	22.351	156	4.639	0.23	141.5	42	7	3.12	0
Drugs	Prolintane	Training	1.36	2.811	21	-4.66	1	1.412	120.3	23.367	3.473	0.462	15.176	147	3.823	0.66	81.2	37	5	2.87	0
Drugs	Propranolol	Training	0.81	2.528	26	-3.73	1	0.820	204.6	23.743	8.863	0.571	6.466	197	4.008	0.88	103.5	12	2	3.28	0
Drugs	Propylhexedrine	Training	-0.30	1.531	13	-3.70	0	1.000	68.7	17.945	0.610	0.363	6.675	1	3.680	0.24	38.7	2	0	2.58	0
Drugs	Sotalol	Training	-1.97	2.492	25	-1.99	1	1.557	173.7	16.544	8.628	0.627	11.586	87	4.598	0.61	125.3	19	8	3.29	0
Drugs	Strychnine	Training	0.21	3.185	51	-4.08	1	0.815	198.6	33.442	3.126	0.610	28.239	3598	4.084	0.27	53.8	331	17	2.72	1
Drugs	Tapentadol	Training	0.28	2.535	21	-4.21	1	0.740	119.2	30.463	1.649	0.528	19.380	75	3.853	0.30	93.4	18	0	2.85	0
Drugs	Testolactone	Training	3.07	2.900	40	-4.63	1	0.648	177.2	42.652	6.837	0.517	20.813	1072	3.956	0.57	65.7	0	0	3.04	0
Drugs	Timolol	Training	-0.71	2.627	30	-2.48	1	0.886	282.7	28.327	8.492	0.527	9.471	277	4.176	0.74	149.8	147	12	3.42	0
Drugs	Tramadol	Training	0.17	2.837	28	-3.52	1	0.879	161.3	38.240	3.511	0.520	20.811	231	3.898	0.41	101.1	24	0	2.93	0
Drugs	Triamterene	Training	1.10	2.624	30	-3.89	1	1.066	133.1	13.338	5.783	0.585	9.246	440	4.035	0.08	75.7	119	16	2.99	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
Drugs	Trichlormethiazide	Training	0.55	2.712	34	-1.67	1	1.136	150.9	10.314	4.550	0.736	5.647	290	4.685	0.06	98.7	63	10	2.92	0
Drugs	Xipamide	Training	1.15	3.077	36	-3.58	1	0.991	203.8	25.099	6.960	0.706	12.075	600	4.628	0.13	164.0	45	4	3.19	0
Explosives	2,6-DA-4-NT	Select	0.92	0.647	17	-2.44	0	0.331	67.2	10.904	0.428	0.676	1.350	0	3.925	0.00	42.8	8	0	2.38	0
Explosives	2,6-DNT	Select	2.10	0.602	18	-3.45	0	0.268	82.6	11.432	0.702	0.610	1.428	12	3.961	0.00	50.9	32	0	2.44	0
Explosives	DPA	Select	3.12	2.197	17	-4.55	0	0.929	75.7	11.478	1.689	0.350	2.835	64	3.850	0.09	43.4	12	4	2.77	0
Explosives	ETN	Select	3.21	2.222	24	-2.39	0	1.092	334.1	28.217	6.072	0.478	8.091	36	3.757	0.24	190.0	70	0	3.04	0
Explosives	NQ	Select	-2.56	0.000	7	-0.45	0	0.877	44.7	0.009	0.003	0.769	1.048	0	3.643	0.00	21.0	0	4	1.96	0
Explosives	PETN	Select	3.64	2.116	26	-2.78	1	0.953	349.9	34.980	4.727	0.459	1.168	0	3.769	0.00	210.0	48	0	2.98	0
Drugs	6-MAM	Select	0.50	3.117	45	-3.54	1	0.814	230.1	34.582	3.711	0.638	27.230	1675	4.071	0.37	77.0	75	2	2.90	1
Drugs	Andarine	Select	3.26	3.156	47	-3.64	1	0.699	505.3	30.983	14.125	0.664	7.121	1039	3.947	0.76	367.6	152	8	3.89	0
Drugs	Atenolol	Select	-2.19	2.361	24	-2.20	1	0.847	232.9	24.089	10.802	0.638	6.117	114	3.837	0.78	143.4	11	2	3.51	0
Drugs	Carvedilol	Select	2.67	3.192	44	-3.41	1	0.959	523.6	40.610	13.141	0.544	12.231	1612	4.164	0.60	246.7	112	5	3.92	0
Drugs	Chlorothiazide	Select	-0.10	2.062	29	-1.50	0	1.357	101.8	9.836	2.301	0.718	3.468	142	4.687	0.12	64.9	31	5	2.63	0
Drugs	Chlorthalidone	Select	0.41	3.074	38	-2.88	1	0.955	163.2	21.909	4.108	0.657	7.151	784	4.628	0.23	118.6	42	6	3.01	0
Drugs	Desacetyl deflazacort	Select	1.84	3.438	55	-3.14	1	0.656	281.3	51.335	8.635	0.578	37.159	3662	4.027	1.33	118.6	92	6	3.23	1
Drugs	Fenfluramine	Select	0.48	2.220	22	-4.92	1	0.616	140.2	18.865	3.396	0.592	15.726	50	3.859	0.12	89.3	4	2	2.99	0
Drugs	Fluticasone prop.	Select	3.71	3.872	61	-4.56	1	0.899	403.2	74.954	9.797	0.570	45.384	4150	4.051	1.96	226.2	0	0	3.34	0
Drugs	Hydrochloroth.	Select	0.00	2.062	29	-0.80	0	1.357	104.1	10.589	2.363	0.714	4.498	142	4.683	0.16	64.9	31	5	2.63	0
Drugs	Isometheptene	Select	-0.03	1.658	10	-3.64	0	1.111	61.5	8.371	1.994	0.493	5.600	0	3.602	0.14	45.0	1	0	2.68	0
Drugs	Mefenorex	Select	0.98	2.313	16	-4.54	1	0.510	109.8	17.102	3.531	0.571	10.762	35	3.805	0.18	71.3	6	2	3.08	0
Drugs	Methylprednisolone	Select	1.97	3.374	50	-2.98	1	0.532	248.4	53.888	8.148	0.546	33.301	2242	3.999	1.20	119.9	0	0	3.16	0
Drugs	Pethidine	Select	1.56	2.687	26	-3.72	1	0.882	148.7	27.102	2.692	0.540	14.312	141	3.890	0.37	101.2	21	1	2.81	0
Drugs	Terbutaline	Select	-1.94	2.322	23	-2.39	1	0.661	134.9	18.749	4.478	0.618	10.540	52	3.856	0.31	91.2	7	3	2.93	0
Drugs	Triamcinolone	Select	0.92	3.282	53	-2.14	1	0.570	269.4	48.288	8.029	0.599	37.658	2577	4.011	1.15	128.5	0	0	3.15	0
Explosives	1,3-DNG	Test	0.97	0.796	13	-1.16	0	1.037	155.4	13.009	4.560	0.625	0.001	0	3.657	0.00	66.0	8	0	2.88	0
Explosives	3-NT	Test	2.45	0.000	13	-3.45	0	0.207	48.3	6.235	0.105	0.616	2.418	0	3.885	0.00	28.4	5	0	2.23	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
Explosives	4-Am-2,6-DNT	Test	1.88	0.602	20	-2.98	0	0.313	92.6	17.731	0.731	0.661	1.289	16	3.976	0.00	60.0	36	0	2.46	0
Explosives	DMDPU	Test	2.55	2.719	25	-4.36	1	0.877	156.1	22.560	5.490	0.464	9.020	232	3.856	0.30	104.7	84	8	3.06	0
Explosives	PYX	Test	7.04	3.890	68	-5.37	0	0.425	688.3	91.684	7.772	0.433	46.089	4107	4.075	0.52	617.0	1198	12	3.76	0
Explosives	RDX	Test	-1.20	1.247	21	-2.65	0	0.667	166.0	5.283	0.481	0.470	15.254	0	3.850	0.00	72.5	114	15	2.56	0
Explosives	TATB	Test	2.93	1.645	27	-1.42	0	0.472	141.3	15.055	0.985	0.446	6.782	147	4.042	0.00	103.7	120	0	2.53	0
Drugs	Amiloride	Test	-0.92	1.906	21	-1.47	0	0.590	109.3	6.614	3.401	0.722	6.940	40	3.962	0.06	75.4	37	12	2.78	0
Drugs	Beclomethasone	Test	2.25	3.428	53	-3.18	1	0.488	257.1	45.846	8.449	0.573	38.015	2577	4.073	1.18	128.5	0	0	3.15	0
Drugs	Betaxolol	Test	0.43	2.771	29	-3.21	1	0.932	337.7	25.957	14.313	0.544	12.070	213	3.836	0.89	194.4	15	2	3.79	0
Drugs	Bisoprolol	Test	-0.22	2.784	28	-2.84	1	1.065	401.2	28.937	13.703	0.514	12.677	192	3.831	0.78	221.1	15	2	3.88	0
Drugs	Carphedon	Test	0.31	2.153	23	-2.12	1	0.684	123.9	16.424	4.182	0.634	8.109	136	3.863	0.10	75.3	44	5	2.95	0
Drugs	Deflazacort	Test	2.43	3.699	59	-3.51	1	0.742	371.5	54.005	11.321	0.562	37.245	4063	4.028	1.39	164.3	104	6	3.42	1
Drugs	Desonide	Test	2.79	3.503	58	-3.04	1	0.711	319.3	64.118	8.391	0.543	45.365	3993	4.021	1.51	130.0	0	0	3.26	1
Drugs	Esmolol	Test	-0.42	2.663	26	-3.06	1	0.942	291.1	28.075	13.131	0.583	9.705	137	3.836	0.86	179.9	13	2	3.68	0
Drugs	Fenetylline	Test	-0.07	2.912	38	-3.64	1	0.958	311.5	24.820	7.170	0.602	27.491	931	3.981	0.40	176.7	344	17	3.58	0
Drugs	Gestrinone	Test	3.26	3.036	41	-5.00	1	0.463	171.3	35.074	5.841	0.572	27.661	1325	3.999	0.47	81.0	0	0	3.09	0
Drugs	Norfenfluramine	Test	0.18	1.650	20	-4.40	0	0.619	104.0	12.083	2.294	0.649	7.545	14	3.857	0.20	64.5	0	0	2.71	0
Drugs	Phenmetrazine	Test	0.09	1.727	18	-2.81	0	0.995	80.5	11.828	0.773	0.592	6.657	56	3.836	0.16	40.3	9	2	2.56	0
Drugs	Pindolol	Test	-0.65	2.411	25	-2.56	1	0.895	193.3	24.184	8.413	0.640	7.334	179	4.042	0.91	97.4	34	4	3.22	0
Drugs	Piretanide	Test	0.07	3.627	39	-2.62	1	1.423	252.4	28.471	4.744	0.647	23.698	737	4.623	0.61	181.3	93	5	3.07	0
Drugs	Salmeterol	Test	1.62	3.045	37	-4.05	0	0.720	585.7	61.428	19.868	0.508	13.376	683	3.872	1.14	364.2	36	1	4.30	0
OGSR	2,4-DNDPA	External	3.76	2.645	27	-4.44	1	0.285	165.4	16.843	5.027	0.464	7.190	226	3.974	0.07	108.9	89	4	3.09	0
OGSR	2-NDPA	External	3.58	2.491	22	-4.39	0	0.285	117.4	10.775	1.444	0.438	7.006	139	3.923	0.10	69.3	49	4	2.86	0
OGSR	4,4'-DNDPA	External	3.78	2.432	27	-4.44	1	0.285	183.7	15.473	7.890	0.464	4.540	236	3.935	0.11	118.5	76	4	3.37	0
OGSR	4-NDPA	External	3.36	2.322	22	-4.39	0	0.285	125.5	13.135	5.219	0.438	3.251	132	3.910	0.10	76.5	38	4	3.11	0
OGSR	Dimethyl phthalate	External	1.88	1.572	18	-3.22	0	0.975	104.7	14.274	0.634	0.588	2.328	11	3.870	0.28	61.4	0	0	2.57	0
OGSR	N-nitroso-DPA	External	3.03	2.293	20	-4.38	0	0.266	104.2	9.676	1.560	0.378	4.490	112	3.874	0.05	63.6	48	8	2.74	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
MEKPs	MEKP CP di.	External	2.08	1.609	18	-3.02	0	0.688	125.4	9.711	0.135	0.376	3.850	2	3.686	0.34	46.0	0	0	2.48	0
MEKPs	MEKP CP tet.	External	4.50	3.425	36	-4.55	0	0.719	440.1	48.528	2.879	0.311	57.758	238	3.703	0.78	153.0	0	0	3.03	0
MEKPs	MEKP CP tri.	External	3.44	2.755	27	-3.83	0	0.708	248.9	31.984	1.803	0.335	32.247	57	3.695	1.02	88.4	0	0	2.64	0
MEKPs	MEKP DHP di.	External	3.19	2.097	18	-2.69	0	0.583	203.9	16.991	0.754	0.454	12.279	5	3.621	0.57	91.0	0	0	2.71	0
MEKPs	MEKP DHP mon.	External	1.33	0.000	9	-1.61	0	0.500	71.6	2.311	0.001	0.585	0.735	0	3.541	0.00	28.0	0	0	1.90	0
MEKPs	MEKP DHP pent.	External	8.28	3.483	45	-4.98	0	0.667	994.0	96.347	9.896	0.334	61.314	386	3.693	1.58	496.0	0	0	3.82	0
MEKPs	MEKP DHP tet.	External	6.65	3.195	36	-4.29	0	0.650	665.8	73.588	6.402	0.358	47.792	178	3.678	1.35	325.0	0	0	3.54	0
MEKPs	MEKP DHP tri.	External	4.91	2.790	27	-3.54	1	0.625	402.3	40.946	3.151	0.394	20.500	51	3.656	0.94	190.0	0	0	3.19	0
MNs	M-1,2,3,4,5-PN	External	3.74	2.932	34	-2.65	1	1.064	479.5	31.980	4.753	0.484	20.233	264	3.809	0.76	351.0	165	0	3.13	0
MNs	M-1,2,3,4,6-PN	External	3.79	2.870	34	-2.65	1	1.064	489.6	34.541	7.361	0.484	13.388	228	3.800	0.70	351.0	147	0	3.24	0
MNs	M-1,2,3,5,6-PN	External	3.74	2.860	34	-2.65	1	1.064	494.2	30.090	7.537	0.484	22.142	255	3.792	0.74	351.0	148	0	3.28	0
MNs	M-1,2,3,4-TN	External	2.28	2.709	30	-1.83	0	1.037	395.1	27.039	4.425	0.521	13.185	135	3.793	0.71	276.0	107	0	3.03	0
MNs	M-1,2,3,5-TN	External	2.35	2.645	30	-1.83	0	1.037	402.5	29.224	5.985	0.521	10.147	174	3.783	0.47	276.0	108	0	3.15	0
MNs	M-1,2,4,5-TN	External	2.60	2.617	30	-1.83	0	1.037	404.3	30.898	4.707	0.521	18.353	187	3.781	0.72	276.0	108	0	3.17	0
MNs	M-1,3,4,5-TN	External	2.57	2.630	30	-1.83	0	1.037	400.7	29.809	4.856	0.521	14.722	162	3.789	0.54	276.0	107	0	3.14	0
MNs	M-2,3,4,5-TN	External	2.32	2.709	30	-1.83	0	1.037	391.0	28.691	2.606	0.521	24.638	187	3.800	0.64	276.0	122	0	3.00	0
MNs	M-1,2,3,6-TN	External	2.32	2.617	30	-1.83	0	1.037	412.2	39.288	9.222	0.521	9.457	138	3.774	0.62	276.0	93	0	3.27	0
MNs	M-1,2,4,6-TN	External	2.60	2.561	30	-1.83	0	1.037	413.9	34.167	7.463	0.521	11.542	159	3.769	0.49	276.0	93	0	3.28	0
MNs	M-1,3,4,6-TN	External	2.73	2.545	30	-1.83	0	1.037	410.3	28.793	7.458	0.521	9.916	141	3.777	0.59	276.0	92	0	3.24	0
MNs	M-1,2,5,6-TN	External	2.28	2.604	30	-1.83	0	1.037	417.5	32.864	9.261	0.521	12.370	169	3.763	0.46	276.0	94	0	3.31	0
MNs	M-1,2,3-TriN	External	0.72	2.439	26	-1.03	0	1.012	317.2	31.283	6.126	0.549	9.562	57	3.767	0.51	210.0	62	0	2.99	0
MNs	M-1,2,4-TriN	External	0.92	2.335	26	-1.03	0	1.012	321.3	25.383	4.486	0.549	10.353	91	3.760	0.40	210.0	62	0	3.07	0
MNs	M-1,3,4-TriN	External	0.87	2.314	26	-1.03	0	1.012	319.1	21.706	4.500	0.549	10.148	75	3.768	0.55	210.0	61	0	3.04	0
MNs	M-2,3,4-TriN	External	0.80	2.421	26	-1.03	0	1.012	310.0	22.729	2.137	0.549	14.953	85	3.781	0.51	210.0	73	0	2.88	0
MNs	M-1,2,6-TriN	External	0.69	2.318	26	-1.03	0	1.012	336.9	30.889	9.400	0.549	9.029	85	3.738	0.39	210.0	51	0	3.29	0
MNs	M-1,3,6-TriN	External	0.91	2.219	26	-1.03	0	1.012	334.7	37.032	9.394	0.549	8.043	84	3.743	0.53	210.0	50	0	3.26	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
SNs	S-2,3,4,5,6-PN	External	3.74	2.932	34	-2.65	1	1.064	479.5	39.612	5.046	0.484	14.254	264	3.809	0.77	351.0	165	0	3.13	0
SNs	S-1,3,4,5,6-PN	External	3.79	2.870	34	-2.65	1	1.064	489.6	38.529	7.847	0.484	11.517	228	3.800	0.68	351.0	147	0	3.24	0
SNs	S-1,2,4,5,6-PN	External	3.74	2.860	34	-2.65	1	1.064	494.2	42.495	7.572	0.484	20.996	255	3.792	0.78	351.0	148	0	3.28	0
SNs	S-1,2,3,5,6-PN	External	3.74	2.860	34	-2.65	1	1.064	494.2	39.796	9.173	0.484	18.856	255	3.792	0.78	351.0	148	0	3.28	0
SNs	S-1,2,3,4,6-PN	External	3.79	2.870	34	-2.65	1	1.064	489.6	37.501	7.874	0.484	18.945	228	3.800	0.81	351.0	147	0	3.24	0
SNs	S-1,2,3,4,5-PN	External	3.74	2.932	34	-2.65	1	1.064	479.5	33.674	4.175	0.484	28.340	264	3.809	0.62	351.0	165	0	3.13	0
SNs	S-3,4,5,6-TN	External	2.28	2.709	30	-1.83	0	1.037	395.1	30.752	5.105	0.521	19.690	135	3.793	0.83	276.0	107	0	3.03	0
SNs	S-2,4,5,6-TN	External	2.35	2.645	30	-1.83	0	1.037	402.5	29.893	6.037	0.521	17.048	174	3.783	0.66	276.0	108	0	3.15	0
SNs	S-2,3,5,6-TN	External	2.60	2.617	30	-1.83	0	1.037	404.3	33.053	4.075	0.521	28.133	187	3.781	0.58	276.0	108	0	3.17	0
SNs	S-2,3,4,6-TN	External	2.57	2.630	30	-1.83	0	1.037	400.7	27.779	5.373	0.521	12.010	162	3.789	0.58	276.0	107	0	3.14	0
SNs	S-2,3,4,5-TN	External	2.32	2.709	30	-1.83	0	1.037	391.0	30.330	2.890	0.521	14.899	187	3.800	0.71	276.0	122	0	3.00	0
SNs	S-1,4,5,6-TN	External	2.32	2.617	30	-1.83	0	1.037	412.2	33.047	9.256	0.521	16.095	138	3.774	0.52	276.0	93	0	3.27	0
SNs	S-1,3,5,6-TN	External	2.60	2.561	30	-1.83	0	1.037	413.9	36.822	7.397	0.521	10.496	159	3.769	0.45	276.0	93	0	3.28	0
SNs	S-1,3,4,6-TN	External	2.73	2.545	30	-1.83	0	1.037	410.3	30.493	7.945	0.521	8.414	141	3.777	0.69	276.0	92	0	3.24	0
SNs	S-1,3,4,5-TN	External	2.57	2.630	30	-1.83	0	1.037	400.7	30.102	5.151	0.521	11.153	162	3.789	0.75	276.0	107	0	3.14	0
SNs	S-1,2,5,6-TN	External	2.28	2.604	30	-1.83	0	1.037	417.5	30.710	7.658	0.521	8.357	169	3.763	0.54	276.0	94	0	3.31	0
SNs	S-1,2,4,6-TN	External	2.60	2.561	30	-1.83	0	1.037	413.9	35.512	7.884	0.521	6.783	159	3.769	0.35	276.0	93	0	3.28	0
SNs	S-1,2,4,5-TN	External	2.60	2.617	30	-1.83	0	1.037	404.3	36.570	5.104	0.521	10.545	187	3.781	0.44	276.0	108	0	3.17	0
SNs	S-1,2,3,6-TN	External	2.32	2.617	30	-1.83	0	1.037	412.2	37.457	9.267	0.521	12.741	138	3.774	0.57	276.0	93	0	3.27	0
SNs	S-1,2,3,5-TN	External	2.35	2.645	30	-1.83	0	1.037	402.5	33.190	6.159	0.521	11.297	174	3.783	0.58	276.0	108	0	3.15	0
SNs	S-1,2,3,4-TN	External	2.28	2.709	30	-1.83	0	1.037	395.1	31.416	5.226	0.521	11.537	135	3.793	0.65	276.0	107	0	3.03	0
SNs	S-1,2,3-TrN	External	0.72	2.439	26	-1.03	0	1.012	317.2	29.677	6.153	0.549	12.741	57	3.767	0.47	210.0	62	0	2.99	0
SNs	S-1,2,4-TrN	External	0.92	2.335	26	-1.03	0	1.012	321.3	28.345	5.324	0.549	6.914	91	3.760	0.36	210.0	62	0	3.07	0
SNs	S-1,2,5-TrN	External	0.87	2.318	26	-1.03	0	1.012	327.8	27.922	6.269	0.549	5.852	113	3.750	0.45	210.0	63	0	3.18	0
SNs	S-1,2,6-TrN	External	0.69	2.318	26	-1.03	0	1.012	336.9	34.262	9.400	0.549	4.423	85	3.738	0.46	210.0	51	0	3.29	0
SNs	S-1,3,4-TrN	External	0.87	2.314	26	-1.03	0	1.012	319.1	24.828	5.335	0.549	8.614	75	3.768	0.65	210.0	61	0	3.04	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
SNs	S-1,3,5-TrN	External	1.03	2.260	26	-1.03	0	1.012	325.6	28.008	4.401	0.549	6.152	105	3.756	0.33	210.0	62	0	3.15	0
SNs	S-1,3,6-TrN	External	0.69	2.318	26	-1.03	0	1.012	336.9	31.193	9.402	0.549	6.903	85	3.738	0.35	210.0	51	0	3.29	0
SNs	S-2,3,4-TrN	External	0.80	2.421	26	-1.03	0	1.012	310.0	20.722	2.257	0.549	12.020	85	3.781	0.53	210.0	73	0	2.88	0
SNs	S-2,3,5-TrN	External	1.05	2.335	26	-1.03	0	1.012	316.6	18.922	4.166	0.549	10.952	122	3.770	0.54	210.0	74	0	3.02	0
SNs	S-2,3,6-TrN	External	0.83	2.297	26	-1.03	0	1.012	325.7	23.187	6.033	0.549	12.855	97	3.759	0.63	210.0	62	0	3.16	0
SNs	S-2,4,5-TrN	External	1.05	2.335	26	-1.03	0	1.012	316.6	26.651	2.310	0.549	9.694	122	3.770	0.32	210.0	74	0	3.02	0
SNs	S-2,4,6-TrN	External	1.03	2.260	26	-1.03	0	1.012	325.6	23.784	4.973	0.549	6.192	105	3.756	0.21	210.0	62	0	3.15	0
SNs	S-2,5,6-TrN	External	0.87	2.318	26	-1.03	0	1.012	327.8	23.464	6.139	0.549	9.725	113	3.750	0.46	210.0	63	0	3.18	0
SNs	S-3,4,5-TrN	External	0.80	2.421	26	-1.03	0	1.012	310.0	22.233	2.863	0.549	12.049	85	3.781	0.73	210.0	73	0	2.88	0
SNs	S-3,4,6-TrN	External	0.87	2.314	26	-1.03	0	1.012	319.1	23.019	5.276	0.549	9.718	75	3.768	0.66	210.0	61	0	3.04	0
SNs	S-3,5,6-TrN	External	0.92	2.335	26	-1.03	0	1.012	321.3	28.640	4.486	0.549	9.418	91	3.760	0.37	210.0	62	0	3.07	0
SNs	S-4,5,6-TrN	External	0.72	2.439	26	-1.03	0	1.012	317.2	27.690	6.141	0.549	15.173	57	3.767	0.46	210.0	62	0	2.99	0
SNs	S-1,2-DN	External	-0.70	2.069	22	-0.28	0	0.992	244.3	26.451	6.297	0.563	4.395	29	3.728	0.37	153.0	29	0	3.00	0
SNs	S-1,3-DN	External	-0.46	1.941	22	-0.28	0	0.992	244.0	31.356	6.294	0.563	7.979	30	3.731	0.30	153.0	28	0	3.00	0
SNs	S-1,4-DN	External	-0.46	1.883	22	-0.28	0	0.992	246.4	22.422	5.461	0.563	4.536	43	3.728	0.36	153.0	28	0	3.05	0
SNs	S-1,5-DN	External	-0.52	1.915	22	-0.28	0	0.992	251.6	24.358	6.308	0.563	4.047	56	3.717	0.31	153.0	29	0	3.15	0
SNs	S-1,6-DN	External	-0.69	1.915	22	-0.28	0	0.992	259.8	30.341	9.533	0.563	2.779	43	3.697	0.35	153.0	20	0	3.26	0
SNs	S-2,3-DN	External	-0.54	2.042	22	-0.28	0	0.992	235.8	15.384	2.645	0.563	12.855	32	3.750	0.50	153.0	37	0	2.83	0
SNs	S-2,4-DN	External	-0.25	1.941	22	-0.28	0	0.992	238.2	16.626	1.912	0.563	6.396	53	3.744	0.18	153.0	37	0	2.89	0
SNs	S-2,5-DN	External	-0.21	1.915	22	-0.28	0	0.992	243.4	15.288	4.263	0.563	6.261	73	3.733	0.31	153.0	38	0	3.02	0
SNs	S-2,6-DN	External	-0.52	1.915	22	-0.28	0	0.992	251.6	21.624	6.142	0.563	4.844	56	3.717	0.39	153.0	29	0	3.15	0
SNs	S-3,4-DN	External	-0.60	2.015	22	-0.28	0	0.992	232.9	15.936	1.960	0.563	9.721	25	3.757	0.59	153.0	36	0	2.73	0
SNs	S-3,5-DN	External	-0.25	1.941	22	-0.28	0	0.992	238.2	20.282	2.501	0.563	7.062	53	3.744	0.34	153.0	37	0	2.89	0
SNs	S-3,6-DN	External	-0.46	1.883	22	-0.28	0	0.992	246.4	27.235	6.329	0.563	6.988	43	3.728	0.27	153.0	28	0	3.05	0
SNs	S-4,5-DN	External	-0.54	2.042	22	-0.28	0	0.992	235.8	23.533	2.321	0.563	8.015	32	3.750	0.43	153.0	37	0	2.83	0
SNs	S-4,6-DN	External	-0.46	1.941	22	-0.28	0	0.992	244.0	22.512	4.762	0.563	5.666	30	3.731	0.35	153.0	28	0	3.00	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4- 3N	IDE	nR08
SNs	S-5,6-DN	External	-0.70	2.069	22	-0.28	0	0.992	244.3	23.954	6.298	0.563	6.901	29	3.728	0.27	153.0	29	0	3.00	0
SNs	S-1-MN	External	-2.18	1.516	18	0.42	0	0.983	173.4	22.530	6.463	0.554	2.751	14	3.675	0.24	105.0	7	0	2.97	0
SNs	S-2-MN	External	-1.79	1.516	18	0.42	0	0.983	166.4	13.813	2.632	0.554	4.816	16	3.702	0.29	105.0	13	0	2.81	0
SNs	S-3-MN	External	-1.98	1.468	18	0.42	0	0.983	163.0	21.458	2.701	0.554	6.375	5	3.713	0.20	105.0	12	0	2.67	0
SNs	S-4-MN	External	-1.98	1.468	18	0.42	0	0.983	163.0	15.226	1.735	0.554	5.795	5	3.713	0.33	105.0	12	0	2.67	0
SNs	S-5-MN	External	-1.79	1.516	18	0.42	0	0.983	166.4	17.138	2.796	0.554	4.045	16	3.702	0.23	105.0	13	0	2.81	0
SNs	S-6-MN	External	-2.18	1.516	18	0.42	0	0.983	173.4	23.123	6.458	0.554	2.777	14	3.675	0.22	105.0	7	0	2.97	0
$C_3H_6O_6N_6$	2,4,6-Trinitro-1,3,5-triazinane	External	-3.18	1.247	21	-2.65	0	0.835	146.8	4.443	0.449	0.530	5.133	0	3.858	0.00	72.5	102	12	2.56	0
$C_3H_6O_6N_6$	N,N',N''-Trinitro-1,2,3-cyclopropanetriamine	External	-4.56	1.484	21	-1.49	0	0.500	176.7	3.581	1.322	0.530	3.682	0	3.901	0.00	89.9	78	12	2.79	0
$C_3H_6O_6N_6$	1,3,5-Trinitro-1,2,3-triazinane	External	-1.27	1.247	21	-2.27	0	0.500	161.0	4.688	0.720	0.592	13.767	0	3.854	0.00	72.5	110	14	2.56	0
$C_3H_6O_6N_6$	2-(N'-Nitro-N-nitroso-carbamimidamido) ethyl nitrate	External	-2.14	1.639	17	-1.56	0	0.833	200.6	7.127	3.062	0.772	7.427	20	3.704	0.09	105.0	50	9	3.03	0
$C_3H_6O_6N_6$	1,2,2-Trinitro-1,3,5-triazinane	External	0.63	0.000	22	-3.04	0	0.779	138.7	3.754	0.028	0.608	4.056	24	3.939	0.00	71.2	155	15	2.27	0
$C_{17}H_7O_{16}N_{11}$	3,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-2,5-pyridinediamine	External	5.78	3.887	68	-5.37	0	0.425	687.0	57.601	11.336	0.433	25.662	4105	4.075	0.87	631.6	1218	12	3.84	0
$C_{17}H_7O_{16}N_{11}$	2,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-3,5-pyridinediamine	External	6.09	3.883	68	-5.37	0	0.425	673.6	68.083	12.839	0.433	15.335	4107	4.075	0.70	617.0	1238	12	3.76	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4-3N	IDE	nR08
$C_6H_6O_6N_6$	4,5,6-Trinitro-1,2,3-benzenetriamine	External	2.48	1.261	27	-1.42	0	0.472	139.8	16.633	0.732	0.446	20.123	123	4.050	0.00	102.1	112	0	2.49	0
$C_6H_6O_6N_6$	(1Z,2Z,3E,4Z,5Z,6E)-N,N',N'',N''',N'''',N''''-Hexahydroxy-1,2,3,4,5,6-cyclohexane hexaimine	External	-4.96	1.372	24	0.55	0	0.472	168.9	15.684	1.485	0.408	6.107	159	3.892	0.09	102.5	126	0	2.59	0
$C_6H_6O_6N_6$	N,N',N''-Trihydroxy-2,4,6-trinitroso- 1,3,5-benzenetriamine	External	-2.56	1.372	24	0.18	0	0.472	165.1	15.592	1.134	0.611	13.058	159	3.950	0.31	102.5	126	0	2.59	0
$C_6H_6O_6N_6$	N,N',N'',N''',N'''',N''''-Hexahydroxy-1,2,3,4,5,6-cyclohexane hexaimine	External	-4.96	1.372	24	0.55	0	0.472	168.9	15.684	1.485	0.408	6.107	159	3.892	0.09	102.5	126	0	2.59	0
$C_6H_6O_6N_6$	1-[(2Z)-2-Methyl-3-nitro-2-propen-1- yl]-3,5-dinitro-1H-1,2,4-triazole	External	0.28	2.162	25	-3.62	0	0.551	181.8	9.464	2.056	0.643	7.953	98	3.997	0.16	118.8	204	13	3.03	0
$C_6H_6O_6N_6$	1-(2-Methyl-3-nitro-2-propen-1-yl)- 3,5-dinitro-1H-1,2,4-triazole	External	0.28	2.162	25	-3.62	0	0.551	181.8	10.901	4.229	0.643	6.465	98	3.997	0.23	118.8	204	13	3.03	0

Table A. 6 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part A

			Molecular descriptors (1-19 out of 32)																		
Dataset	Analyte	ANN subset	LogD (pH=7)	ATS 6p	MPC 02	BLTD 48	Hyper tens-80	GATS 1m	Aeigv	RDF 050e	DP15	BIC1	RDF 055e	p4p4	BEH m1	H7u	D/D	p1p4 -4N	p4-3N	IDE	nR08
$C_6H_6O_6N_6$	(1E,2E,3E,4E,5E,6E)-N,N',N'',N''',N'''', N''''''-Hexahydroxy-1,2,3,4,5,6-cyclohexane hexaimine	External	-4.96	1.372	24	0.55	0	0.472	168.9	15.684	1.485	0.408	6.107	159	3.892	0.09	102.5	126	0	2.59	0
$C_6H_6O_6N_6$	N-{1-[Amino(oxo)acetyl]-3,5-dioxo- 1,2,4-triazolidin-4-yl} ethanediamide	External	-6.27	2.002	26	1.06	0	1.261	201.7	8.280	5.892	0.535	6.253	61	3.790	0.00	118.1	153	24	2.99	0

Table A. 7: Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)													Best ANN, MLP:32-2-1			Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-06 8	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2 -6N	H-04 8	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	1,2-DNB	Training	-1.436	27.8	0	0.028	1.60	1.433	1	0.000	45.8	0.00	0	106	0	7.96	8.69	0.73	8.68	0.72
Explosives	1,2-DNG	Training	-1.457	32.7	0	0.025	0.86	0.724	2	0.000	44.8	0.02	0	92	0	4.42	4.29	-0.13	4.13	-0.29
Explosives	1,3-DNB	Training	-1.546	27.2	0	0.020	1.55	1.356	3	0.000	39.5	0.01	0	102	0	9.04	9.43	0.39	9.14	0.10
Explosives	1-MNG	Training	0.000	18.9	0	0.018	-0.43	0.387	1	0.000	21.0	0.01	0	9	0	1.85	2.15	0.30	2.04	0.19
Explosives	2,4-DA-6-NT	Training	-1.514	24.4	0	0.020	0.92	1.366	3	0.000	73.1	0.00	0	50	0	3.66	5.06	1.40	5.21	1.55
Explosives	2,4-DNT	Training	-1.362	32.9	0	0.022	2.10	1.415	5	0.000	50.7	0.01	0	138	0	11.21	10.08	-1.13	9.45	-1.76
Explosives	2-Am-4,6-DNT	Training	-1.024	41.1	0	0.020	1.88	1.352	7	0.000	74.7	0.01	0	178	0	10.67	8.43	-2.24	8.22	-2.45
Explosives	2-NT	Training	0.000	14.1	0	0.027	2.45	1.393	0	0.000	45.8	0.00	0	25	0	10.29	9.17	-1.12	9.18	-1.11
Explosives	3,4-DNT	Training	-1.436	33.0	0	0.021	2.15	1.482	4	0.000	50.9	0.06	0	140	0	10.13	9.86	-0.27	9.39	-0.74
Explosives	3,5-DNA	Training	-1.469	34.2	0	0.014	1.63	1.314	5	0.000	68.5	0.01	0	136	0	9.26	8.01	-1.25	7.97	-1.29
Explosives	4-NT	Training	0.000	13.4	0	0.021	2.45	1.403	2	0.000	28.8	0.00	0	23	0	10.47	10.04	-0.43	9.76	-0.71
Explosives	DADP	Training	0.000	13.5	0	0.018	1.23	0.815	16	0.000	46.8	0.04	0	0	0	8.43	7.80	-0.63	7.61	-0.82
Explosives	DEDPU	Training	0.108	30.7	0	0.025	3.26	3.172	3	0.094	167.5	0.18	1	512	0	14.81	13.50	-1.31	12.84	-1.97
Explosives	DMNB	Training	-1.649	65.4	0	0.029	1.82	0.551	4	0.000	77.9	0.00	0	48	0	7.30	9.24	1.94	9.15	1.85
Explosives	EGDN	Training	0.000	16.3	0	0.020	1.51	0.549	1	0.000	19.3	0.00	0	38	0	5.44	5.34	-0.10	5.57	0.13
Explosives	HMDD	Training	-1.108	10.0	2	0.029	0.00	1.390	1	0.000	71.2	0.01	0	120	8	4.43	2.72	-1.71	2.65	-1.78
Explosives	HMTA	Training	-1.600	5.7	4	0.018	0.99	1.607	0	0.000	65.3	0.00	0	156	12	1.52	2.65	1.13	2.57	1.05
Explosives	HMTD	Training	-1.108	13.4	2	0.025	0.13	1.842	1	0.000	87.7	0.03	0	210	12	2.80	3.57	0.77	3.49	0.69
Explosives	HMX	Training	1.000	100.2	4	0.019	-1.71	1.602	20	0.250	105.8	0.19	0	2068	8	4.64	3.07	-1.57	3.86	-0.78
Explosives	HNDPA	Training	1.610	187.6	0	0.018	4.44	2.479	70	1.035	182.7	0.18	0	6229	0	15.57	16.28	0.71	16.11	0.54
Explosives	MHN	Training	2.035	277.9	0	0.012	5.16	1.794	36	0.969	244.4	0.27	0	4958	0	14.55	16.09	1.54	15.53	0.98
Explosives	NB	Training	0.000	10.7	0	0.025	1.81	1.340	0	0.000	25.8	0.00	0	15	0	8.00	9.30	1.30	9.38	1.38
Explosives	NG	Training	0.771	50.2	0	0.015	2.41	0.562	5	0.125	55.6	0.02	0	326	0	8.97	8.92	-0.05	8.82	-0.16

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Explosives	NM	Training	0.000	2.0	0	0.035	0.10	0.178	0	0.000	5.9	0.00	0	0	0	1.85	1.69	-0.16	2.45	0.60
Explosives	PA	Training	-0.408	63.8	0	0.018	1.75	1.279	11	0.000	81.9	0.01	0	424	0	4.96	5.32	0.36	6.20	1.24
Explosives	PGDN	Training	0.000	24.8	0	0.030	1.87	0.661	2	0.000	34.4	0.02	0	60	0	7.50	6.34	-1.16	6.52	-0.99
Explosives	R-Salt	Training	-1.427	22.2	0	0.027	-1.18	1.060	0	0.000	61.3	0.00	0	189	6	3.54	2.57	-0.97	2.76	-0.79
Explosives	SHN	Training	2.035	277.9	0	0.012	5.16	1.536	36	0.969	236.2	0.28	0	4958	0	14.55	16.15	1.60	15.83	1.28
Explosives	TATP	Training	-1.000	33.2	0	0.012	2.16	0.968	48	0.000	183.3	0.10	0	0	0	11.99	12.44	0.45	11.19	-0.80
Explosives	TEGDN	Training	0.484	47.2	0	0.020	1.46	0.823	1	0.094	111.1	0.00	0	284	0	7.61	9.19	1.58	8.76	1.15
Explosives	Tetryl	Training	1.000	102.6	0	0.029	1.25	2.130	21	0.125	102.6	0.12	0	1358	0	11.83	10.66	-1.17	10.81	-1.02
Explosives	TMETN	Training	0.771	78.0	0	0.016	3.08	0.491	15	0.375	115.1	0.07	0	483	0	11.78	10.90	-0.88	11.15	-0.63
Explosives	TNB	Training	-0.408	53.0	0	0.014	1.36	1.315	9	0.000	81.0	0.01	0	342	0	9.81	10.06	0.25	9.98	0.16
Explosives	TNT	Training	-0.408	62.2	0	0.019	1.79	1.359	11	0.000	84.0	0.02	0	424	0	12.25	10.46	-1.79	10.18	-2.07
Drugs	Acebutolol	Training	0.580	91.8	0	0.016	1.68	1.905	16	0.365	272.0	0.20	1	279	0	6.38	6.82	0.44	7.27	0.89
Drugs	Acetazolamide	Training	-1.173	28.5	0	0.039	-0.31	0.320	9	0.000	26.8	0.03	0	49	0	2.15	2.53	0.38	2.25	0.10
Drugs	AICAR	Training	-0.218	47.8	0	0.018	-2.49	1.970	14	0.247	99.5	0.08	0	252	1	1.42	2.60	1.18	1.86	0.44
Drugs	Alprenolol	Training	-0.667	39.2	0	0.018	2.78	2.267	5	0.165	144.2	0.15	0	63	0	11.41	7.61	-3.80	7.81	-3.60
Drugs	Altizide	Training	0.265	74.8	0	0.036	1.10	0.384	44	0.358	218.3	0.06	1	212	1	6.87	4.98	-1.89	5.55	-1.32
Drugs	Amino-glutethimide	Training	-0.505	38.1	0	0.021	1.27	2.060	12	0.062	107.1	0.06	0	51	0	4.99	5.44	0.44	6.15	1.16
Drugs	Anastrazole	Training	0.077	49.6	0	0.024	2.68	2.749	42	0.339	292.8	0.08	1	653	0	8.02	11.77	3.75	10.57	2.55
Drugs	Bambuterol	Training	0.727	101.5	0	0.012	1.50	2.998	52	0.786	555.1	0.22	1	1325	0	8.41	7.26	-1.15	6.61	-1.80
Drugs	Bendro-flumethiazide	Training	0.764	97.7	0	0.024	1.49	2.272	105	0.491	210.8	0.17	0	418	1	8.50	7.23	-1.27	7.37	-1.13
Drugs	Benzoylecgonine	Training	0.531	46.9	1	0.024	2.29	2.431	15	0.278	126.5	0.10	1	362	0	4.05	5.90	1.85	6.18	2.13
Drugs	Benzthiazide	Training	1.000	73.1	0	0.030	1.98	1.582	57	0.408	213.9	0.08	0	383	0	9.39	8.16	-1.23	8.73	-0.66

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-06 8	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2 -6N	H-04 8	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Beta\dexameth.	Training	1.361	110.4	0	0.017	1.92	2.909	143	0.596	231.7	0.38	0	0	0	10.89	11.37	0.48	11.03	0.14
Drugs	Budesonide	Training	1.425	91.6	0	0.017	3.02	3.427	123	0.729	387.3	0.37	0	0	1	15.29	13.09	-2.20	12.76	-2.53
Drugs	Bumetanide	Training	0.706	76.5	0	0.026	3.00	3.168	25	0.530	343.8	0.16	1	183	0	6.25	6.96	0.71	7.19	0.94
Drugs	Bumetanide	Training	0.706	76.5	0	0.026	3.00	3.168	25	0.530	343.8	0.16	1	183	0	6.28	6.96	0.68	7.19	0.91
Drugs	Bunolol	Training	-0.087	50.0	0	0.017	2.57	2.038	25	0.293	139.2	0.15	1	99	0	7.59	7.39	-0.20	7.76	0.17
Drugs	Bupropion	Training	-1.000	37.8	0	0.031	3.08	1.594	15	0.207	96.6	0.08	0	25	0	9.54	10.96	1.42	10.76	1.22
Drugs	Butizide	Training	-0.151	71.8	0	0.035	0.79	1.147	56	0.299	144.5	0.11	1	181	1	6.73	5.32	-1.41	5.60	-1.13
Drugs	Canrenone	Training	0.988	49.9	0	0.018	2.50	3.212	91	0.411	182.9	0.25	1	0	0	14.71	13.69	-1.02	12.72	-1.99
Drugs	Carteolol	Training	-0.005	47.2	0	0.018	1.84	2.247	26	0.255	135.6	0.16	1	199	0	4.17	5.27	1.10	5.96	1.79
Drugs	Celiprolol	Training	0.744	113.8	0	0.015	2.06	2.549	43	0.564	318.1	0.26	0	932	0	7.66	8.10	0.44	9.38	1.72
Drugs	Chlorexolone	Training	0.142	49.9	0	0.033	1.49	1.774	31	0.293	107.7	0.09	1	327	0	9.74	8.81	-0.93	8.37	-1.37
Drugs	Clobenzorex	Training	-0.502	23.1	0	0.027	4.54	2.966	4	0.177	171.4	0.10	0	75	0	12.70	13.95	1.25	13.50	0.80
Drugs	Clobetasol	Training	1.361	108.7	0	0.017	2.74	2.751	143	0.596	226.5	0.41	0	0	0	12.94	12.98	0.04	12.61	-0.33
Drugs	Clopamide	Training	-0.031	64.0	1	0.033	1.75	1.351	35	0.330	139.0	0.12	1	441	0	6.01	8.85	2.84	8.46	2.45
Drugs	Cyclopenthiazide	Training	0.265	66.7	0	0.033	1.52	1.440	56	0.373	217.4	0.09	1	272	1	9.61	7.33	-2.29	7.77	-1.85
Drugs	Danazol	Training	0.796	46.7	0	0.021	4.33	3.285	90	0.402	210.7	0.21	0	259	1	16.65	15.83	-0.82	15.54	-1.11
Drugs	Diacetolol	Training	0.355	75.5	0	0.014	1.00	1.711	16	0.259	154.1	0.22	1	214	0	3.62	4.98	1.36	5.50	1.88
Drugs	Dichlorphenamide	Training	-1.000	60.9	0	0.070	0.90	0.099	38	0.000	76.4	0.03	0	0	0	4.19	5.05	0.86	4.85	0.66
Drugs	Efaproxiral	Training	0.528	64.1	0	0.017	3.83	3.249	37	0.306	200.3	0.17	0	163	0	7.28	9.98	2.70	10.58	3.30
Drugs	Epitizide	Training	0.510	101.8	0	0.035	1.03	0.288	96	0.583	206.3	0.05	0	270	1	6.68	5.15	-1.53	5.60	-1.08
Drugs	Etacr. acid	Training	0.228	63.8	0	0.071	3.38	1.777	12	0.177	91.9	0.19	0	0	0	7.31	5.61	-1.71	6.98	-0.33
Drugs	Etamivan	Training	-0.778	36.4	0	0.021	1.46	1.674	5	0.181	125.5	0.10	0	109	0	5.75	6.10	0.35	6.49	0.74
Drugs	Famprofazone	Training	1.000	51.5	2	0.022	3.76	4.649	26	0.440	354.6	0.37	0	2245	0	17.86	15.83	-2.03	15.63	-2.23
Drugs	Fenbutrazate	Training	1.000	52.3	1	0.019	4.48	4.153	13	0.215	373.3	0.22	0	699	0	17.61	15.07	-2.54	15.13	-2.48

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Fencamfamin	Training	-0.954	18.0	0	0.022	3.44	2.868	1	0.043	178.8	0.10	0	62	0	10.22	7.98	-2.24	8.43	-1.79
Drugs	Fludrocortisone	Training	1.208	105.4	0	0.017	1.54	2.642	130	0.505	229.8	0.32	0	0	0	9.59	10.32	0.73	10.53	0.94
Drugs	Flumethasone	Training	1.418	128.1	0	0.017	1.74	2.983	161	0.658	264.7	0.36	0	0	0	11.13	11.46	0.33	10.78	-0.35
Drugs	Formoterol	Training	0.807	61.3	0	0.022	1.61	3.061	17	0.321	353.5	0.15	1	368	0	7.71	6.81	-0.90	7.26	-0.45
Drugs	FPCAM	Training	1.427	145.8	0	0.016	2.58	3.084	182	0.842	370.0	0.35	0	0	0	10.79	9.49	-1.30	9.72	-1.07
Drugs	Fulvestrant	Training	1.859	149.3	0	0.017	7.56	2.737	153	0.652	535.4	0.51	0	0	0	18.08	16.76	-1.32	17.80	-0.28
Drugs	Furosemide	Training	0.414	59.0	0	0.031	2.35	1.677	26	0.265	113.5	0.07	1	114	1	3.36	3.99	0.63	4.56	1.20
Drugs	Hydroflumeth.	Training	0.214	76.1	0	0.026	0.24	0.630	77	0.167	102.3	0.11	0	175	2	2.63	2.87	0.24	2.97	0.34
Drugs	Labetalol	Training	0.484	63.9	0	0.026	2.74	3.199	16	0.224	166.9	0.16	1	165	0	9.84	10.05	0.21	9.83	-0.02
Drugs	Mefruside	Training	0.121	83.6	1	0.030	1.67	1.154	81	0.163	205.6	0.09	1	402	0	9.61	7.20	-2.41	6.72	-2.89
Drugs	Metoprolol	Training	-0.437	42.0	0	0.019	1.85	2.201	6	0.224	105.8	0.17	1	75	0	7.27	6.72	-0.55	6.55	-0.72
Drugs	Nadoxolol	Training	-0.111	34.2	0	0.019	1.65	2.440	6	0.200	109.0	0.06	1	99	0	9.26	8.06	-1.20	7.42	-1.84
Drugs	Nikethamide	Training	-1.344	19.8	0	0.028	0.54	1.797	0	0.056	75.9	0.07	0	75	0	3.61	6.02	2.41	5.92	2.31
Drugs	Ostarine	Training	1.286	140.8	0	0.015	3.26	3.182	91	0.776	197.7	0.23	0	1567	0	13.34	14.21	0.87	12.97	-0.37
Drugs	Oxprenolol	Training	-0.522	42.6	0	0.015	2.29	1.749	5	0.193	174.4	0.14	1	75	0	9.46	6.68	-2.78	6.46	-3.00
Drugs	Oxycodone	Training	0.829	47.1	1	0.024	0.91	2.054	50	0.171	228.0	0.07	1	589	0	4.60	4.29	-0.31	4.70	0.10
Drugs	Oxymorphone	Training	0.712	48.2	1	0.018	1.01	2.140	50	0.087	210.8	0.05	1	529	0	2.94	4.18	1.24	4.47	1.53
Drugs	Pemoline	Training	-1.000	15.7	0	0.031	0.50	2.174	1	0.028	49.0	0.01	0	19	0	3.15	4.53	1.38	4.52	1.37
Drugs	Pentazocine	Training	0.082	33.3	1	0.023	3.80	1.494	28	0.311	199.7	0.16	0	353	0	8.59	10.54	1.95	11.34	2.75
Drugs	Pentetrazol	Training	-1.915	7.8	0	0.028	0.22	1.355	0	0.000	50.4	0.00	0	58	0	2.86	4.91	2.05	5.17	2.31
Drugs	Phendimetrazine	Training	-1.085	15.4	1	0.030	1.91	2.125	1	0.028	91.5	0.07	0	102	0	6.32	7.46	1.14	7.10	0.78
Drugs	Polythiazide	Training	0.816	110.2	1	0.035	1.59	0.414	107	0.583	276.5	0.08	0	628	1	8.84	6.34	-2.50	6.32	-2.52
Drugs	Prenylamine	Training	1.000	33.8	0	0.020	5.40	4.189	6	0.241	329.9	0.13	0	204	0	15.88	15.75	-0.13	15.83	-0.05
Drugs	Probenecid	Training	-0.287	69.2	1	0.018	3.05	1.424	12	0.212	189.5	0.05	1	183	0	5.30	6.26	0.96	6.69	1.39

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-06 8	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2 -6N	H-04 8	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Prolintane	Training	-1.000	18.6	1	0.023	4.35	2.100	1	0.087	137.2	0.14	0	144	0	9.25	9.20	-0.05	9.24	-0.01
Drugs	Propranolol	Training	-0.618	30.8	0	0.020	3.26	2.480	7	0.168	107.9	0.17	1	86	0	11.61	9.53	-2.08	9.07	-2.54
Drugs	Propylhexedrine	Training	-2.000	13.6	0	0.031	2.72	0.938	0	0.028	82.8	0.06	0	13	0	8.76	7.62	-1.14	8.28	-0.48
Drugs	Sotalol	Training	-0.768	54.0	0	0.016	0.29	2.106	18	0.247	91.0	0.17	1	102	0	2.91	3.53	0.62	3.77	0.86
Drugs	Strychnine	Training	1.116	35.1	1	0.019	1.62	4.210	51	0.102	247.0	0.12	1	1535	0	8.14	8.25	0.11	7.64	-0.50
Drugs	Tapentadol	Training	-0.864	30.7	1	0.021	2.64	1.799	5	0.128	191.3	0.10	0	163	0	7.33	7.18	-0.15	7.82	0.49
Drugs	Testolactone	Training	0.655	43.3	0	0.020	3.07	2.442	53	0.255	167.0	0.16	1	0	0	9.17	12.34	3.17	12.00	2.83
Drugs	Timolol	Training	-0.275	39.7	0	0.019	1.53	1.721	19	0.240	145.4	0.12	1	650	0	7.95	7.13	-0.82	7.53	-0.42
Drugs	Tramadol	Training	-0.362	38.1	1	0.024	2.54	2.022	11	0.172	179.7	0.11	1	308	0	7.12	6.47	-0.65	7.07	-0.05
Drugs	Triamterene	Training	-0.345	22.7	0	0.034	1.20	2.535	11	0.137	95.0	0.03	0	348	0	6.56	6.34	-0.22	7.09	0.53
Drugs	Trichlor-methiazide	Training	0.075	67.8	0	0.036	0.76	-0.150	56	0.233	98.6	0.06	0	156	2	4.92	3.84	-1.08	4.67	-0.25
Drugs	Xipamide	Training	0.106	74.0	0	0.030	2.88	1.788	43	0.352	172.1	0.03	1	130	0	5.01	8.36	3.35	8.06	3.05
Explosives	2,6-DA-4-NT	Select	-1.545	24.0	0	0.020	0.92	1.330	4	0.000	64.6	0.01	0	48	0	3.16	5.64	2.48	5.67	2.51
Explosives	2,6-DNT	Select	-1.312	33.7	0	0.021	2.10	1.400	3	0.000	59.6	0.00	0	140	0	11.04	9.63	-1.41	9.20	-1.84
Explosives	DPA	Select	-1.175	9.7	0	0.036	3.12	2.087	0	0.049	52.5	0.02	0	22	0	13.97	13.27	-0.70	12.91	-1.06
Explosives	ETN	Select	1.000	103.9	0	0.013	3.21	0.836	12	0.312	101.0	0.12	0	1092	0	11.60	12.84	1.24	12.27	0.67
Explosives	NQ	Select	0.000	8.5	0	0.029	-0.98	0.420	0	0.000	13.2	0.00	0	0	0	1.49	0.90	-0.59	0.24	-1.25
Explosives	PETN	Select	1.000	119.4	0	0.016	3.64	0.518	22	0.750	196.0	0.08	0	1248	0	12.39	10.48	-1.91	10.75	-1.64
Drugs	6-MAM	Select	0.836	42.2	1	0.017	1.44	2.667	46	0.390	198.9	0.11	1	653	0	4.45	6.23	1.78	6.25	1.80
Drugs	Andarine	Select	1.286	140.8	0	0.015	3.26	3.182	91	0.776	197.7	0.23	0	1567	0	12.40	14.21	1.81	12.97	0.57
Drugs	Atenolol	Select	-0.332	50.0	0	0.019	0.24	2.134	9	0.227	89.5	0.18	1	74	0	2.73	3.70	0.97	3.84	1.11
Drugs	Carvedilol	Select	1.271	47.0	0	0.028	4.06	3.971	18	0.333	328.3	0.23	0	688	0	13.30	13.77	0.47	13.73	0.43
Drugs	Chlorothiazide	Select	-0.867	46.8	0	0.039	-0.02	0.527	36	0.083	71.2	0.08	0	106	0	2.08	3.85	1.77	3.89	1.81

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Chlorthalidone	Select	0.514	59.4	0	0.032	0.41	2.765	52	0.575	165.5	0.05	1	114	0	4.91	5.33	0.42	5.49	0.58
Drugs	Desacetyl deflazacort	Select	1.425	81.1	0	0.016	1.84	3.630	123	0.623	273.1	0.38	0	288	0	10.85	11.31	0.46	11.05	0.20
Drugs	Fenfluramine	Select	-0.492	42.0	0	0.026	3.34	1.898	9	0.137	124.4	0.03	0	39	0	9.29	8.09	-1.20	8.64	-0.65
Drugs	Fluticasone prop.	Select	1.790	164.2	0	0.016	3.71	2.407	182	0.979	360.3	0.52	0	0	2	16.21	15.33	-0.88	14.55	-1.66
Drugs	Hydrochloroth.	Select	-0.869	47.9	0	0.039	0.01	0.322	36	0.083	75.5	0.07	0	106	2	2.13	3.55	1.42	3.72	1.59
Drugs	Isometheptene	Select	0.000	15.8	0	0.033	2.99	0.641	1	0.000	64.8	0.05	0	5	0	6.99	6.74	-0.25	7.29	0.30
Drugs	Mefenorex	Select	-1.000	19.3	0	0.030	3.51	1.787	0	0.043	134.7	0.05	0	27	0	8.50	9.04	0.54	9.27	0.77
Drugs	Methyl-prednisolone	Select	1.119	89.2	0	0.018	1.97	2.563	111	0.536	250.1	0.35	0	0	0	11.44	11.02	-0.42	11.42	-0.02
Drugs	Pethidine	Select	-0.569	33.5	1	0.021	2.45	1.969	6	0.149	217.3	0.08	0	242	0	8.41	8.01	-0.40	8.09	-0.32
Drugs	Terbutaline	Select	-1.000	44.7	0	0.020	0.52	1.649	18	0.300	106.3	0.07	0	32	0	2.83	3.40	0.57	3.73	0.90
Drugs	Triamcinolone	Select	1.398	113.2	0	0.017	0.92	2.849	143	0.596	241.6	0.33	0	0	0	7.24	8.37	1.13	8.58	1.34
Explosives	1,3-DNG	Test	-1.476	32.1	0	0.013	0.97	0.431	3	0.125	24.4	0.03	0	90	0	4.02	4.40	0.38	4.47	0.45
Explosives	3-NT	Test	0.000	13.6	0	0.026	2.45	1.345	1	0.000	39.4	0.00	0	23	0	10.90	9.11	-1.79	9.21	-1.69
Explosives	4-Am-2,6-DNT	Test	-1.104	41.5	0	0.019	1.88	1.352	6	0.000	82.5	0.01	0	180	0	10.30	7.94	-2.36	7.96	-2.34
Explosives	DMDPU	Test	-0.610	24.4	0	0.022	2.55	2.301	3	0.063	109.2	0.14	0	412	0	12.40	11.51	-0.89	11.29	-1.11
Explosives	PYX	Test	2.591	291.0	0	0.013	7.04	4.372	160	1.608	468.9	0.29	0	21069	0	16.90	16.78	-0.12	18.08	1.18
Explosives	RDX	Test	-0.222	51.9	3	0.021	-1.20	1.120	9	0.000	88.4	0.06	0	597	6	5.95	4.00	-1.95	4.04	-1.91
Explosives	TATB	Test	0.619	80.1	0	0.014	2.93	2.017	15	0.000	122.8	0.17	0	615	0	9.40	11.72	2.32	11.92	2.52
Drugs	Amiloride	Test	-0.991	29.4	0	0.036	1.58	1.231	8	0.118	50.2	0.06	0	78	0	3.63	3.63	0.00	4.53	0.90
Drugs	Beclomethasone	Test	1.369	101.1	0	0.018	2.25	2.431	143	0.596	233.1	0.42	0	0	0	11.50	12.45	0.95	12.03	0.53
Drugs	Betaxolol	Test	0.265	43.1	0	0.016	2.87	2.161	10	0.259	144.3	0.22	1	131	0	11.60	10.21	-1.39	9.67	-1.93
Drugs	Bisoprolol	Test	0.512	61.9	0	0.016	2.21	2.620	10	0.302	210.2	0.23	1	132	0	9.59	8.68	-0.91	8.43	-1.16

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)													Best ANN, MLP:32-2-1			Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
Drugs	Carphedon	Test	-0.505	29.7	0	0.029	0.31	2.145	5	0.134	81.7	0.04	0	140	0	5.16	4.10	-1.06	4.65	-0.51
Drugs	Deflazacort	Test	1.857	93.8	0	0.015	2.43	3.762	144	0.896	315.5	0.44	0	365	0	13.60	13.50	-0.10	12.71	-0.89
Drugs	Desonide	Test	1.425	88.5	0	0.016	2.79	3.097	175	0.707	305.1	0.41	0	0	0	12.50	12.94	0.44	12.19	-0.31
Drugs	Esmolol	Test	0.252	58.0	0	0.016	2.02	2.016	10	0.299	108.1	0.21	1	101	0	8.36	7.31	-1.05	7.40	-0.96
Drugs	Fenetylline	Test	1.013	49.2	0	0.025	2.20	3.974	20	0.293	275.9	0.12	1	2120	0	9.46	9.68	0.22	9.53	0.06
Drugs	Gestrinone	Test	0.673	48.0	0	0.019	3.26	2.784	46	0.365	193.9	0.20	0	0	1	14.00	13.14	-0.86	13.15	-0.85
Drugs	Norfenfluramine	Test	-1.000	35.6	0	0.027	2.78	1.803	9	0.000	72.4	0.05	0	0	0	8.60	8.17	-0.43	8.01	-0.59
Drugs	Phenmetrazine	Test	-1.177	13.4	0	0.029	1.67	1.959	0	0.000	75.9	0.04	0	26	0	5.19	5.43	0.24	5.68	0.49
Drugs	Pindolol	Test	-0.656	28.1	0	0.024	1.86	2.188	7	0.184	95.1	0.17	0	148	1	4.86	5.13	0.27	5.92	1.06
Drugs	Piretanide	Test	0.744	62.0	0	0.022	2.80	3.527	33	0.513	316.4	0.15	1	567	0	5.12	7.46	2.34	7.40	2.28
Drugs	Salmeterol	Test	1.000	86.0	0	0.017	3.90	3.131	7	0.149	173.1	0.36	0	320	0	15.60	15.43	-0.17	15.61	0.01
OGSR	2,4-DNDPA	External	0.424	43.3	0	0.030	3.76	2.976	9	0.240	86.9	0.04	0	650	0	16.58	13.94	-2.64	13.87	-2.71
OGSR	2-NDPA	External	-0.766	23.2	0	0.032	3.58	2.937	2	0.115	80.0	0.02	0	225	0	16.43	14.76	-1.67	14.37	-2.06
OGSR	4,4'-DNDPA	External	0.401	38.9	0	0.028	3.78	2.516	12	0.299	73.1	0.04	0	647	0	14.87	14.05	-0.82	13.97	-0.90
OGSR	4-NDPA	External	-0.950	21.8	0	0.035	3.36	2.281	4	0.174	60.6	0.03	0	222	0	14.19	14.53	0.34	14.22	0.03
OGSR	Dimethyl phthalate	External	-0.869	27.4	0	0.037	1.88	1.506	1	0.000	118.1	0.03	0	0	0	7.38	7.91	0.53	7.50	0.12
OGSR	N-nitroso-DPA	External	-1.000	16.2	0	0.035	3.03	2.292	0	0.049	69.7	0.01	0	156	0	13.45	14.27	0.82	14.04	0.59
MEKPs	MEKP CP di.	External	-1.486	20.9	0	0.030	2.08	0.989	16	0.000	82.1	0.03	0	0	0	-	11.16	-	10.29	-
MEKPs	MEKP CP tet.	External	0.437	94.2	0	0.022	4.50	2.479	96	0.535	412.0	0.12	0	0	0	-	16.37	-	16.14	-
MEKPs	MEKP CP tri.	External	-0.673	50.7	0	0.027	3.44	0.998	48	0.000	246.8	0.12	0	0	0	-	15.54	-	14.72	-
MEKPs	MEKP DHP di.	External	-1.345	58.5	0	0.026	3.19	0.568	16	0.000	121.4	0.05	0	0	0	-	13.28	-	12.11	-
MEKPs	MEKP DHP mon.	External	0.000	16.9	0	0.037	1.33	0.416	0	0.000	51.1	0.01	0	0	0	-	4.15	-	4.79	-
MEKPs	MEKP DHP pent.	External	0.740	337.8	0	0.017	8.28	0.831	160	0.781	528.2	0.28	0	0	0	-	16.76	-	17.64	-

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
MEKPs	MEKP DHP tet.	External	0.425	218.6	0	0.017	6.65	0.696	96	0.531	375.7	0.24	0	0	0	-	16.63	-	17.19	-
MEKPs	MEKP DHP tri.	External	-0.432	125.6	0	0.021	4.91	1.107	48	0.281	308.1	0.14	0	0	0	-	15.20	-	14.62	-
MNs	M-1,2,3,4,5-PN	External	1.720	227.41	0	0.017	3.74	1.598	27	0.719	226.7	0.22	0	3143	0	-	14.25	-	13.06	-
MNs	M-1,2,3,4,6-PN	External	1.894	223.7	0	0.012	3.79	1.467	28	0.750	210.7	0.22	0	3141	0	-	14.06	-	13.03	-
MNs	M-1,2,3,5,6-PN	External	1.894	222.5	0	0.013	3.74	1.712	28	0.656	184.6	0.21	0	3141	0	-	14.53	-	13.33	-
MNs	M-1,2,3,4-TN	External	1.366	180.5	0	0.013	2.28	1.391	20	0.500	191.2	0.20	0	1830	0	-	12.40	-	11.39	-
MNs	M-1,2,3,5-TN	External	1.502	179.5	0	0.017	2.35	1.256	20	0.531	157.6	0.17	0	1830	0	-	11.87	-	11.26	-
MNs	M-1,2,4,5-TN	External	1.390	179.3	0	0.017	2.60	1.239	20	0.469	178.0	0.18	0	1830	0	-	12.76	-	12.01	-
MNs	M-1,3,4,5-TN	External	1.442	180.1	0	0.018	2.57	1.108	20	0.469	191.4	0.17	0	1830	0	-	12.35	-	11.67	-
MNs	M-2,3,4,5-TN	External	1.472	183.0	0	0.018	2.32	1.264	19	0.437	218.9	0.17	0	1832	0	-	12.67	-	11.72	-
MNs	M-1,2,3,6-TN	External	1.464	176.5	0	0.012	2.32	1.137	21	0.438	123.8	0.22	0	1828	0	-	11.97	-	11.66	-
MNs	M-1,2,4,6-TN	External	1.361	176.4	0	0.015	2.60	1.377	21	0.563	157.2	0.16	0	1828	0	-	12.14	-	11.69	-
MNs	M-1,3,4,6-TN	External	1.390	177.2	0	0.013	2.73	0.994	21	0.594	177.0	0.17	0	1828	0	-	12.69	-	11.91	-
MNs	M-1,2,5,6-TN	External	1.353	175.5	0	0.013	2.28	1.348	21	0.344	126.8	0.17	0	1828	0	-	12.24	-	11.53	-
MNs	M-1,2,3-TriN	External	1.000	139.9	0	0.013	0.72	1.008	14	0.312	120.3	0.18	0	940	0	-	7.33	-	7.63	-
MNs	M-1,2,4-TriN	External	1.000	139.7	0	0.015	0.92	1.235	14	0.375	149.1	0.13	0	940	0	-	7.33	-	7.52	-
MNs	M-1,3,4-TriN	External	1.000	140.1	0	0.013	0.87	0.930	14	0.312	156.0	0.15	0	940	0	-	7.80	-	7.65	-
MNs	M-2,3,4-TriN	External	1.000	142.4	0	0.018	0.80	1.052	13	0.187	183.7	0.14	0	942	0	-	7.72	-	7.53	-
MNs	M-1,2,6-TriN	External	1.000	136.7	0	0.014	0.69	1.188	15	0.250	90.1	0.15	0	938	0	-	7.27	-	7.63	-
MNs	M-1,3,6-TriN	External	1.000	137.2	0	0.013	0.91	1.245	15	0.406	87.1	0.18	0	938	0	-	7.21	-	7.96	-
SNs	S-2,3,4,5,6-PN	External	1.720	227.4	0	0.016	3.74	1.330	27	0.719	224.9	0.24	0	3143	0	-	13.81	-	12.90	-
SNs	S-1,3,4,5,6-PN	External	1.894	223.7	0	0.014	3.79	1.290	28	0.750	204.7	0.24	0	3141	0	-	13.84	-	13.00	-
SNs	S-1,2,4,5,6-PN	External	1.894	222.5	0	0.014	3.74	1.405	28	0.656	180.5	0.24	0	3141	0	-	14.1	-	13.36	-
SNs	S-1,2,3,5,6-PN	External	1.894	222.5	0	0.012	3.74	1.755	28	0.656	158.5	0.28	0	3141	0	-	14.27	-	13.42	-

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)													Av. t_R^M (min)	Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-06 8	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2 -6N	H-04 8		t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
SNs	S-1,2,3,4,6-PN	External	1.894	223.7	0	0.012	3.79	1.442	28	0.750	202.1	0.28	0	3141	0	-	14.29	-	13.40	-
SNs	S-1,2,3,4,5-PN	External	1.720	227.4	0	0.017	3.74	1.764	27	0.719	225.7	0.20	0	3143	0	-	14.29	-	13.21	-
SNs	S-3,4,5,6-TN	External	1.366	180.5	0	0.013	2.28	1.357	20	0.500	184.3	0.26	0	1830	0	-	12.77	-	11.89	-
SNs	S-2,4,5,6-TN	External	1.502	179.5	0	0.015	2.35	1.622	20	0.531	156.6	0.24	0	1830	0	-	12.46	-	11.77	-
SNs	S-2,3,5,6-TN	External	1.390	179.3	0	0.016	2.60	1.404	20	0.469	172.0	0.16	0	1830	0	-	12.87	-	12.23	-
SNs	S-2,3,4,6-TN	External	1.442	180.1	0	0.020	2.57	1.258	20	0.469	181.7	0.21	0	1830	0	-	12.34	-	11.61	-
SNs	S-2,3,4,5-TN	External	1.472	183.0	0	0.019	2.32	1.333	19	0.437	219.9	0.22	0	1832	0	-	12.13	-	11.40	-
SNs	S-1,4,5,6-TN	External	1.464	176.5	0	0.012	2.32	1.354	21	0.438	126.5	0.22	0	1828	0	-	12.48	-	11.84	-
SNs	S-1,3,5,6-TN	External	1.361	176.4	0	0.015	2.60	1.472	21	0.563	156.0	0.15	0	1828	0	-	11.86	-	11.59	-
SNs	S-1,3,4,6-TN	External	1.390	177.2	0	0.013	2.73	1.172	21	0.594	166.6	0.22	0	1828	0	-	12.66	-	11.98	-
SNs	S-1,3,4,5-TN	External	1.442	180.1	0	0.016	2.57	1.182	20	0.469	193.5	0.22	0	1830	0	-	12.49	-	11.76	-
SNs	S-1,2,5,6-TN	External	1.353	175.5	0	0.017	2.28	1.300	21	0.344	154.8	0.20	0	1828	0	-	11.87	-	11.23	-
SNs	S-1,2,4,6-TN	External	1.361	176.4	0	0.016	2.60	0.975	21	0.563	154.6	0.16	0	1828	0	-	11.66	-	11.42	-
SNs	S-1,2,4,5-TN	External	1.390	179.3	0	0.016	2.60	1.012	20	0.469	172.5	0.16	0	1830	0	-	11.78	-	11.50	-
SNs	S-1,2,3,6-TN	External	1.464	176.5	0	0.013	2.32	1.377	21	0.438	124.3	0.25	0	1828	0	-	12.09	-	11.75	-
SNs	S-1,2,3,5-TN	External	1.502	179.5	0	0.016	2.35	1.347	20	0.531	150.9	0.21	0	1830	0	-	11.91	-	11.45	-
SNs	S-1,2,3,4-TN	External	1.366	180.5	0	0.014	2.28	1.241	20	0.500	181.6	0.22	0	1830	0	-	12.1	-	11.36	-
SNs	S-1,2,3-TrN	External	1.000	139.9	0	0.012	0.72	1.245	14	0.312	119.8	0.21	0	940	0	-	7.51	-	7.76	-
SNs	S-1,2,4-TrN	External	1.000	139.7	0	0.016	0.92	0.920	14	0.375	134.4	0.14	0	940	0	-	7.03	-	7.42	-
SNs	S-1,2,5-TrN	External	1.000	138.9	0	0.017	0.87	0.886	14	0.281	116.1	0.17	0	940	0	-	7.14	-	7.52	-
SNs	S-1,2,6-TrN	External	1.000	136.7	0	0.015	0.69	0.909	15	0.250	84.1	0.18	0	938	0	-	6.87	-	7.51	-
SNs	S-1,3,4-TrN	External	1.000	140.1	0	0.013	0.87	1.088	14	0.312	146.9	0.20	0	940	0	-	7.76	-	7.77	-
SNs	S-1,3,5-TrN	External	1.000	139.5	0	0.017	1.03	0.891	14	0.406	156.0	0.12	0	940	0	-	6.82	-	7.35	-
SNs	S-1,3,6-TrN	External	1.000	136.7	0	0.015	0.69	1.218	15	0.250	89.0	0.15	0	938	0	-	7.02	-	7.48	-

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)													Av. t_R^M (min)	Best ANN, MLP:32-2-1 t_R^P (min)	Best ANN, MLP:32-2-1 Δt_R (min)	Ensemble, 6xMLP:32-2-1 t_R^P (min)	Ensemble, 6xMLP:32-2-1 Δt_R (min)
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-06 8	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2 -6N	H-04 8					
SNs	S-2,3,4-TrN	External	1.000	142.4	0	0.021	0.80	1.220	13	0.187	176.8	0.19	0	942	0	-	7.54	-	7.39	-
SNs	S-2,3,5-TrN	External	1.000	141.7	0	0.018	1.05	1.244	13	0.312	146.0	0.18	0	942	0	-	8.15	-	7.89	-
SNs	S-2,3,6-TrN	External	1.000	139.4	0	0.017	0.83	1.291	14	0.219	125.9	0.22	0	940	0	-	7.94	-	7.93	-
SNs	S-2,4,5-TrN	External	1.000	141.7	0	0.021	1.05	1.000	13	0.312	175.5	0.12	0	942	0	-	6.97	-	7.36	-
SNs	S-2,4,6-TrN	External	1.000	139.5	0	0.022	1.03	0.933	14	0.406	141.8	0.12	0	940	0	-	6.82	-	7.21	-
SNs	S-2,5,6-TrN	External	1.000	138.9	0	0.017	0.87	1.188	14	0.281	121.7	0.16	0	940	0	-	7.60	-	7.67	-
SNs	S-3,4,5-TrN	External	1.000	142.4	0	0.016	0.80	1.119	13	0.187	186.2	0.21	0	942	0	-	8.06	-	7.74	-
SNs	S-3,4,6-TrN	External	1.000	140.1	0	0.013	0.87	1.092	14	0.312	145.4	0.21	0	940	0	-	7.93	-	7.83	-
SNs	S-3,5,6-TrN	External	1.000	139.7	0	0.015	0.92	1.302	14	0.375	144.6	0.13	0	940	0	-	7.03	-	7.43	-
SNs	S-4,5,6-TrN	External	1.000	139.9	0	0.013	0.72	1.187	14	0.312	122.2	0.18	0	940	0	-	7.78	-	7.84	-
SNs	S-1,2-DN	External	0.898	105.7	0	0.015	-0.70	0.759	9	0.187	81.0	0.15	0	392	0	-	3.90	-	4.37	-
SNs	S-1,3-DN	External	0.825	106.0	0	0.013	-0.46	1.082	9	0.250	82.6	0.14	0	392	0	-	3.93	-	4.61	-
SNs	S-1,4-DN	External	0.750	105.9	0	0.015	-0.46	0.797	9	0.187	105.4	0.14	0	392	0	-	4.04	-	4.44	-
SNs	S-1,5-DN	External	0.707	105.4	0	0.019	-0.52	0.748	9	0.156	86.3	0.11	0	392	0	-	3.90	-	4.34	-
SNs	S-1,6-DN	External	0.605	103.7	0	0.014	-0.69	0.792	10	0.219	53.0	0.13	0	390	0	-	3.85	-	4.39	-
SNs	S-2,3-DN	External	0.915	107.7	0	0.018	-0.54	1.133	8	0.062	120.1	0.19	0	394	0	-	4.50	-	4.69	-
SNs	S-2,4-DN	External	0.828	107.6	0	0.022	-0.25	0.888	8	0.187	132.9	0.10	0	394	0	-	3.99	-	4.27	-
SNs	S-2,5-DN	External	0.750	107.1	0	0.018	-0.21	0.765	8	0.187	110.5	0.12	0	394	0	-	4.43	-	4.65	-
SNs	S-2,6-DN	External	0.707	105.4	0	0.020	-0.52	0.806	9	0.156	85.7	0.14	0	392	0	-	4.03	-	4.41	-
SNs	S-3,4-DN	External	0.926	108.0	0	0.012	-0.60	1.010	8	0.000	138.9	0.18	0	394	0	-	4.51	-	4.61	-
SNs	S-3,5-DN	External	0.828	107.6	0	0.018	-0.25	0.830	8	0.187	141.5	0.11	0	394	0	-	4.17	-	4.52	-
SNs	S-3,6-DN	External	0.750	105.9	0	0.015	-0.46	1.076	9	0.187	82.8	0.16	0	392	0	-	3.95	-	4.53	-
SNs	S-4,5-DN	External	0.915	107.7	0	0.017	-0.54	0.822	8	0.062	143.0	0.12	0	394	0	-	4.04	-	4.41	-
SNs	S-4,6-DN	External	0.825	106.0	0	0.014	-0.46	0.776	9	0.250	110.7	0.13	0	392	0	-	4.04	-	4.45	-

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
SNs	S-5,6-DN	External	0.898	105.7	0	0.015	-0.70	1.071	9	0.187	86.1	0.12	0	392	0	-	3.97	-	4.32	-
SNs	S-1-MN	External	-0.817	77.5	0	0.015	-2.18	0.629	5	0.125	50.3	0.10	0	105	0	-	2.53	-	1.91	-
SNs	S-2-MN	External	-0.861	78.7	0	0.021	-1.79	0.633	4	0.062	81.0	0.10	0	107	0	-	2.66	-	2.08	-
SNs	S-3-MN	External	-0.823	79.0	0	0.015	-1.98	0.881	4	0.000	79.9	0.12	0	107	0	-	2.56	-	2.01	-
SNs	S-4-MN	External	-0.823	79.0	0	0.013	-1.98	0.705	4	0.000	100.3	0.11	0	107	0	-	2.66	-	2.04	-
SNs	S-5-MN	External	-0.861	78.7	0	0.021	-1.79	0.570	4	0.062	81.5	0.08	0	107	0	-	2.59	-	2.02	-
SNs	S-6-MN	External	-0.817	77.5	0	0.016	-2.18	0.633	5	0.125	50.3	0.09	0	105	0	-	2.50	-	1.88	-
$C_3H_6O_6N_6$	2,4,6-Trinitro-1,3,5-triazinane	External	-0.136	52.6	0	0.019	1.32	1.064	9	0.000	71.9	0.02	0	426	0	-	3.40	-	4.15	-
$C_3H_6O_6N_6$	N,N',N''-Trinitro-1,2,3-Cyclopropane triamine	External	-0.737	46.9	0	0.018	-0.81	0.552	9	0.000	107.1	0.03	0	465	0	-	2.73	-	2.77	-
$C_3H_6O_6N_6$	1,3,5-Trinitro-1,2,3-triazinane	External	-0.474	50.7	2	0.018	-0.72	1.154	9	0.000	80.0	0.03	0	540	0	-	4.06	-	3.90	-
$C_3H_6O_6N_6$	2-(N'-Nitro-N-nitroso-carbamimidamido) ethyl nitrate	External	-0.367	63.0	0	0.023	-0.15	1.138	4	0.062	58.0	0.03	0	341	0	-	2.78	-	2.64	-
$C_3H_6O_6N_6$	1,2,2-Trinitro-1,3,5-triazinane	External	-0.445	50.7	1	0.024	0.63	1.159	3	0.000	70.8	0.00	0	461	4	-	5.01	-	5.23	-

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2 -6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
$C_{17}H_7O_{16}N_{11}$	3,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-2,5-pyridinediamine	External	2.589	291.0	0	0.016	5.78	4.474	160	1.795	373.8	0.36	0	21070	0	-	16.78	-	17.89	-
$C_{17}H_7O_{16}N_{11}$	2,6-Dinitro-N,N'-bis(2,4,6-trinitrophenyl)-3,5-pyridinediamine	External	2.591	290.6	0	0.014	6.09	3.596	160	1.608	243.8	0.34	0	21071	0	-	16.78	-	17.95	-
$C_6H_6O_6N_6$	4,5,6-Trinitro-1,2,3-benzenetriamine	External	0.619	79.8	0	0.014	2.48	2.157	15	0.000	122.5	0.06	0	615	0	-	11.43	-	11.55	-
$C_6H_6O_6N_6$	(1Z,2Z,3E,4Z,5Z,6E)-,N',N'',N''',N'''',N'''''-hexahydroxy-1,2,3,4,5,6-cyclohexane hexaimine	External	0.963	67.3	0	0.013	0.49	0.847	3	0.000	154.4	0.03	0	450	0	-	2.67	-	3.30	-
$C_6H_6O_6N_6$	N,N',N''-Trihydroxy-2,4,6-trinitroso-1,3,5-benzenetriamine	External	0.530	74.2	0	0.014	-0.63	1.483	3	0.000	134.2	0.08	0	450	0	-	3.32	-	3.39	-

Table A. 7 (Continued): Molecular descriptors and predicted retention times for best ANN and ensemble following forward selection - Part B

			Molecular descriptors (20-32 out of 32)														Best ANN, MLP:32-2-1		Ensemble, 6xMLP:32-2-1	
Dataset	Analyte	ANN subset	Eeig 11d	TIE	N-068	R3v+	LogP	Mor 07p	p2B p2B	GGI7	QXXe	H6v	Neo pla stic-50	p1p1p2-6N	H-048	Av. t_R^M (min)	t_R^P (min)	Δt_R (min)	t_R^P (min)	Δt_R (min)
$C_6H_6O_6N_6$	N,N',N'',N''',N'''',N'''''-Hexahydroxy-1,2,3,4,5,6-Cyclohexane hexaimine	External	0.963	67.3	0	0.013	0.49	0.847	3	0.000	154.4	0.03	0	450	0	-	2.67	-	3.30	-
$C_6H_6O_6N_6$	1-[(2Z)-2-Methyl-3-nitro-2-propen-1-yl]-3,5-dinitro-1H-1,2,4-triazole	External	0.314	61.7	0	0.022	0.28	1.884	13	0.269	111.1	0.06	0	925	1	-	8.75	-	8.75	-
$C_6H_6O_6N_6$	1-(2-Methyl-3-nitro-2-propen-1-yl)-3,5-dinitro-1H-1,2,4-triazole	External	0.314	61.7	0	0.024	0.28	1.695	13	0.269	91.0	0.05	0	925	1	-	9.08	-	8.93	-
$C_6H_6O_6N_6$	(1E,2E,3E,4E,5E,6E)-,N',N'',N''',N'''',N'''''-hexahydroxy-1,2,3,4,5,6-cyclohexane hexaimine	External	0.963	67.3	0	0.013	0.49	0.847	3	0.000	154.4	0.03	0	450	0	-	2.67	-	3.30	-
$C_6H_6O_6N_6$	N-{1-[Amino(oxo)acetyl]-3,5-dioxo-1,2,4-triazolidin-4-yl}ethanediamide	External	0.206	73.7	0	0.019	-4.24	1.688	16	0.187	46.4	0.07	0	402	0	-	1.88	-	0.13	-